New Developments in the NMR-Linked Metabolomics Analysis of Lysosomal Storage Disorders

Benita Claire Percival

1st Supervisor:

Professor Martin Grootveld, Leicester School of Pharmacy, De Montfort University

2nd Supervisor:

Dr Mark Edgar, Department of Chemistry, Loughborough University

JANUARY 16, 2020

Leicester School of Pharmacy, De Montfort University
Thesis for the award of PhD in Metabolomics

Table of Contents

Publications

1. Percival, B. C., Wann, A., Masania, J., Sinclair, J., Sullo, N., Grootveld, M. Detection and determination of methanol and further potential toxins in human saliva collected from cigarette smokers: a ^{1}H NMR investigation. *JSM Biotechnology and Biomedical Engineering.* 2018 5(1): 1081, 1-8. ISSN: 2333-7117.

2. Grootveld, M., Percival B. C., Grootveld, K. Chronic non-communicable disease risks presented by lipid oxidation products in fried foods. *Hepatobiliary Surgery and Nutrition* 2018; 7(4): 305-312. doi: 0.21037/hbsn.2018.04.01. 1-8.

3. Moumtaz S, Percival BC, Parmar D, Grootveld KL, Jansson P, Grootveld M. Generation of toxic α,β-unsaturated and saturated aldehydes during simulated shallow frying episodes: comparisons of common frying oils with a novel high-stability algae oil product. *Nature Scientific Reports* 2019; 9(Article number 4125): 1-21.

4. Edgar M, Percival BC, Gibson M, Beresford K, Masania J, Wilson P, Grootveld M. Benchtop NMR spectroscopy and spectral analysis of the cis- and trans-stilbene products of the Wittig reaction. *Journal of Chemical Education* 2019.

5. Percival BC, Grootveld M, Gibson M, Osman Y, Molinari M, Jafari F, Sahota T, Martin M, Casanova F, Mather ML, Edgar M, Masania J, Wilson PB. Low-Field, Benchtop NMR spectroscopy as a potential tool for point-of-care diagnostics of metabolic conditions: Validation, protocols and computational models. *High Throughput* 2018; 8(1). pii: E2. DOI: 10.3390/ht8010002

6. Grootveld M, Percival BC, Gibson M, Osman Y, Edgar M, Molinari M, Mather ML, Casanova F, Wilson PB. Progress in low-field benchtop NMR spectroscopy in chemical and biochemical Analysis - A review. *Analytica Chimica Acta* 2019-02-23, DOI: 10.1016/j.aca.2019.02.026. Impact Factor 5.123

7. Percival B, Savel E, Ampem G, Gibson M, Edgar M, Jafari F, Frederick K, Wilson P, Grootveld M. Molecular composition of and potential health benefits offered by natural East African virgin sunflower oil products: A 400 MHz ^{1}H NMR analysis study. *International Journal of Nutrition* 2019; 3(3): 22-43. DOI: 10.14302/issn.2379-7835.ijn-19-2677

8. Grootveld M, Percival BC, Moumtaz S, Grootveld KL. A ^{1}H NMR-linked PCR modelling strategy for tracking the fatty acid sources of aldehydic lipid oxidation products in culinary oils exposed to simulated shallow-frying episodes. *World Academy of Science, Engineering and Technology Journal* 2019.

9. Le Gresley A D L, Ampem G, Grootveld M, Percival B, Naughton D. Characterisation of peroxidation products arising from culinary oils exposed to continuous and discontinuous thermal degradation processes *Food and Function* 2019 10 7952-7966.

10. Percival B, Gibson M, Wilson P, Platt FM, Grootveld M. Metabolomic Studies of Lipid Storage Disorders, with Special Reference to Niemann-Pick Type C Disease: A Critical Review with Future Perspectives. *International Journal of Molecular Sciences* 2020 21 (7) 2533

Acknowledgements

I would like to acknowledge Prof. Martin Grootveld for his support throughout the entire PhD process both on a professional and on a personal level and his wife, Kerry. A special thanks to Martin, for teaching me, if about nothing else, about the Queen's English and continuously mocking my Leicester phrases and accent – what it is right is, you have made me laugh so much over the last few years. I would like to thank all the funding bodies who have supported me throughout my PhD programme including De Montfort University for the fee scholarship and the National Tay-Sachs and Allied Diseases (NTSAD) for my bursary. I would like to thank Innovate UK and Wet Engineering for the post as research assistant which supported me throughout my degree. I would like to thank Dr Mark Edgar for teaching me technical skills on NMR at Loughborough University and for all the hilarious NMR-based banter. I would like to acknowledge Dr Philippe Wilson for his support throughout the final years of my PhD. I would like to thank Prof. Tim Claridge for the use of the NMR facility at Oxford University. I would like to thank Dr Victor Ruiz Rodado, Dr Nikol Sullo, Dr Justine Leenders, Dr Jinit Masania, Dr Lucy Owen, Angela Wann, Krishan Chauhan, and Fereshteh Jafari for all the support they have given me as colleagues throughout the four years and the collaborative work completed. A big thanks to everyone passing through the laboratory aka 'the village', project students, ERASMUS placement students and graduate champions, for putting up with me and my banter over this period.

I would like to thank all my volunteers who helped me throughout this process, as without them, this work would not be possible. I would like to thank my Mum, Sheila Ross, and Dad, Jason Ross, for supporting me throughout my PhD, when I kept feeling like giving up, and not being intelligent enough. You truly are the best parents in the world and if I did not have your support, I would not have got through my first degree, let alone my second. I would like to thank my dearest friends, in particular, Hattie Bass, Naomi Clarkson, Megan Robinson, Emily Beesley and Katy Woodason who have stuck by me, despite me being unsociable as ever and trying to help, but not having degrees to do so – but you all did, in so many other ways. Miles Gibson – thank you for correcting my English and taking the time to read through this, you are just not as good as Martin. Also, thank you Nathan Dare, despite being the biggest liability in my life over the past few years, you have supported me through every tear, delirium, high and low - I will always remember that and love you for it.

I would like to dedicate the work to my grandpa, Narendra Chauhan.

Declaration

This thesis is entirely the work of my own and is an original piece of work, unless otherwise indicated, it has not been previously accepted in any application for a degree.

Patient Recruitment and Sample collection

Samples were collected by Krishan Chauhan (Leicester School of Pharmacy, De Montfort University, UK). Yvonne Latour (National Institute of Health, USA) Dr. Kerri-Lee Wallom (Department of Pharmacology, Oxford University, UK) and Allison Bradbury (School of Veterinary Medicine, University of Pennsylvania, Philadelphia, USA).

All plasma healthy control samples were collected by Krishan Chauhan at Leicester School of Pharmacy, De Montfort University, Leicester, UK. Yvonne Latour under Prof. Cynthia Tifft collected all the GM1 and GM2 samples (CSF, plasma and urine) at the National Institute of Health (NIH) clinical centre in Bethesda, Maryland, USA.

Dr. Kerri-Lee Wallom collected the Gaucher disease samples (human urine and plasma).

Dr. Allison Bradbury under Prof. Charles H. Vite collected the feline urine for the Niemann Pick studies.

Sample Preparation

Dr. Victor Ruiz-Rodado and Devki Parmar contributed to sample preparation and acquisition of urine and plasma samples at Oxford University in June 2016 towards the Gaucher's disease research. Furthermore, Devki Parmar aided in the sample preparation of feline urine for the Niemann Pick Disease in November 2016.

Dr Jinit Masania and Lucy Owen provided data from the LC-MS for the Tay Sachs, Sandhoff Disease and GM1 Type II Analysis.

Thesis Outline

Chapter 1: Literature Review

A literature review of lysosomal storage disorders (LSD's) is detailed and the current research exploring the pathogenesis of the diseases, diagnosis and potential treatments, with a specific focus on metabolomics applications is presented. The dysfunction of the lysosome is detrimental to health, characteristically causing serious damage to the central nervous system (CNS), although other major organs can be affected, resulting in short life expectancy. For the purpose of this thesis the review focuses on three lysosomal storage disorders; Gaucher's Disease (GD), Niemann Pick Disease (NPC), GM1 and GM2 Gangliosidosis, are described in-depth and the challenges currently faced with these diseases are described. The lack of metabolomic applications in this field is addressed, which could contribute to further understanding of the disease pathology, clinical prognosis and potential treatments.

Chapter 2: Methods

An in-depth comprehensive methodological review of the mechanisms and history of metabolomics in conjunction with utilising the core techniques nuclear magnetic resonance (NMR) and mass spectrometry (MS). The techniques' compatibility with biofluid analysis is explored, alongside the advantages, disadvantages, operation and theory of both techniques. The evolution of metabolomics is reviewed; expressing the exponential growth in the field in the last few decades, inclusive of applications to blood, urine and cerebrospinal fluid (CSF). Multivariate statistical methods, used to investigate these large datasets acquired by the techniques described above, and the significance of these approaches' methodologies are outlined. The viability of these methods to seek biomarkers and understand disease severity are also addressed and the potential challenges faced in undertaking this study e.g. xenobiotic interactions.

Chapter 3: Gaucher's Disease

Metabolomic data derived from aged-matched plasma samples from human GD patients and corresponding controls aged 35-70 years were acquired using 700 MHz ^{1}H NMR (Oxford University) and analysed *via* multivariate statistics. Principal component analysis (PCA) showed a significant difference between the controls and GD patients and a sub-clustering within the GD cohort. Lipid disturbances are evident in plasma profiles with a reduction in high density lipoprotein (HDL), low density lipoprotein (LDL) and very-low

density lipoprotein (vLDL). Urine poses more of a challenge comparatively due to significantly more drug metabolites present, in addition to naturally occurring metabolites. Drug metabolites were removed and no significant differences were observed in the urine samples, which may be a result of all the GD patients receiving treatment.

Chapter 4: Niemann Pick Disease

^{1}H NMR is used to study feline urine to achieve further understanding into the potential use of hydroxypropyl-β-cyclodextrin (HPβCD) as a treatment to the disorder. Utilising 400MHz ^{1}H NMR (De Montfort University), HPβCD was not detected in feline urine, nor the antibiotics taken by the felines. Three HPβCD treatment approaches were used, two intrathecal methods were administered using a 120 mg dose, one initiated at 3 weeks, pre-symptomatically (n = 36) and another initiated at 6 weeks, post-symptomatically (n = 36). The third approach involved subcutaneous injection of 1,000 mg/kg and 120 mg intrathecally and this was initiated at 3 weeks of age (n = 23). Significant differences were not present upon comparison of the treated and untreated NPC felines using the intrathecal 120 mg dose presymptomatically or post symptomatically. However, the felines treated by subcutaneous injection of 1000 mg/kg and intrathecal injection of 120 mg of HPβCD initiated at 3 weeks of age significant differences were observed. Partial least Square-Discriminant Analysis (PLS-DA) provided a $Q^2 = 0.53$ and $p = 0.028$, utilising permutation and validation methodologies. Significant metabolites included felinine, creatine/creatinine, hippurate, phenylacetylglycine, trimethylamine oxide, betaine and taurine, amongst others which pathways are explored in depth.

Chapter 5: GM1 Type II

^{1}H NMR was used to seek and identify new biomarkers of disease activity, severity and progression in plasma, cerebrospinal fluid and urine from human patients with GM1 Type II gangliosidosis. Urine and plasma were collected from control participants. Statistically significant metabolites include urinary upregulated citrate, N-acetyls, creatinine, creatine and lactate amongst other metabolites. Additionally, in plasma downregulated LDL and upregulated creatinine and lactate amongst other metabolic disturbances were observed. However, these results were interpreted with caution, due to the different ages and fasting status in the two respective cohorts. The impact of xenobiotics was also considered, such as drugs and supplements, that the GM1 Type II cohort were taking as part of this investigation. This study reveals the complexity of analysing lysosomal storage disorders using a human model using a global untargeted approach.

Abbreviations

Word	Abbreviation
1-Methylnicotinaminde	1-MN
2-Hydroxypropyl-β-Cyclodextrin	HPβCD
3-(Trimethylsilyl) propionic-2,2,3,3-d4 acid	TSP
3-D-Hydroxybutryate	3-HB
Adenosine Triphosphate	ATP
Alanine Amino Transferase	ALT
Amoxicillin	AMOX
Analysis of Variance	ANOVA
Area under curve receiver operating characteristic	AUROC
Avance I	AVI
Azithromycin	AZI
Bile Acid	BA
Blood Brain Barrier	BBB
Biological Magnetic Resonance Bank	BMRB
Bis(monoacylglycero)phosphate	BMP
Body Mass Index	BMI
Branched-chain amino acid	BCAA
Cardiovascular Disease	CVD
Carr-Purcell-Meiboom-Gill	CPMG
Central Nervous System	CNS
Cerebrospinal Fluid	CSF
Chaperone Therapy	CT
Clavamox	CLX
Krabbe Disease	KD
Clavulanic acid	CLAV
Confidence Interval	CI
Constant Sum Normalisation	CSN
Correlation Spectroscopy	COSY

Cyclodextrin	CD
De Montfort University	DMU
Deoxyribonucleic Acid	DNA
Dimethyl Sulphoxide	DMSO
Dimethylamine	DMA
Dimethylglycine	DMG
Electron Spray Ionisation Mass Spectrometry	ESI-MS
Electronic Reference To access In Vivo Concentrations	ERETIC
Endoplasmic Reticulum	ER
Enzyme Replacement Therapy	ERT
Ethylenediaminetetraacetic Acid	EDTA
False Discovery Rate	FDR
Fluorescence Detection	FD
Liquid chromatography-electrospray ionisation-mass spectrometry	LC-ESI-MS
Fourier Transform	FT
Free Induction Decay	FID
Galactosylceramide	GalCer
Gangliosidosis GM1	GM1
Gangliosidosis GM2	GM2
Gas Chromatography-Mass Spectrometry	GC-MS
Gastrosesophageal Reflux Disease	GERD
Gaucher Disease	GD
Glucosylceramide	GlcCer
Glycosphingolipids	GSL
Haematopoietic Stem Cell Therapy	HSCT
Heteronuclear Single Quantum Correlation	HSQC
High Density Lipoprotein	HDL
High Density Lipoprotein Cholesterol	HDLc
High Performance Liquid Chromatography	HPLC
Human Metabolome Database	HMDB
Indole-3-lactate	I-3-L

Indoxyl Sulphate	IS
Inductively Coupled Plasma Mass Spectrometry	ICP-MS
Intelligent Selected Bucket	ISB
International Collaborative Gaucher Group	ICGG
International Study of Macro/Mincronutrients and Blood Pressure	INTERMAP
Intrathecal Intravenous	IT
L-acetyl-carnitine	LAC
Leave-one-out cross validation	LOOCV
Limit of Detection	LOD
Lipid Storage Disorders	LLSDs / lipidoses
Liquid Chromatography	LC
Liquid Chromatography – Mass Spectrometry	LC-MS
Low Density Lipoprotein	LDL
Low Density Lipoprotein Cholesterol	LDLc
Lysosomal Storage Disorder	LSD
Mass Spectrometry	MS
Metronidazole	MET
Multivariate	MV
Multivariate Analysis	MVA
N-acetyl group	NAC
National Institute of Health	NIH
Niemann Pick Disease	NP
Niemann Pick Disease Type A	NPA
Niemann Pick Disease Type B	NPB
Niemann Pick Disease Type C	NPC
Nuclear Magnetic Resonance	NMR
Nuclear Overhauser Effect Presaturation	NOEPR
Nuclear Overhauser Effect Spectroscopy	NOESY
Ortho-Partial Least Squares-Discriminant Analysis	O-PLS-DA
Out-of-bag	OOB
Parkinson's Disease	PD

Partial Least Squares-Discriminant Analysis	PLS-DA
Phenyl-acetyl-glycine	PAG
Phosphocreatine	p-Creatine
Presaturation	PRESAT
Principal Component Analysis	PCA
Random Forest	RF
Receiver Operating Curve	ROC
Sandhoff Disease	SD
Selective serotonin reuptake inhibitors	SSRI
Solid Phase Extraction	SPE
Standard Error of the Mean	SEM
Subcutaneous Injection	SQ
Substrate Reduction Therapy	SRT
Support Vector Machine	SVM
Symptomatic and Supportive Therapy	S/S
Tay Sachs Disease	TSD
Thin Layer Chromatography	TLC
Total Correlated Spectroscopy	TOCSY
Triacylglycerol	TAG
Tricarboxylic acid	TCA
Trimethylamine	TMA
Trimethylamine-N-Oxide	TMAO
Ultra performance Liquid Chromatography	UPLC
Ultra Violet – Visible Spectroscopy	UV-Vis
United Kingdom	UK
United States Department of Agriculture	USDA
Variable Importance Plot	VIP
Variables	Var.
Very Low-Density Lipoprotein	vLDL
Gaucher Disease Type 1	GD1

Gaucher Disease Type 3	GD3
2,4(S)-hydroxycholesterol	2,4(S)-HC

List of Figures

determined a p value of $p = 0.2445$ showing overfitting as a result of small sample numbers. (d) PCA scores plot of control samples only grouped by age 20 years or under, 30 years or under and 70 years or younger in red, green and blue respectively. PC1 and PC2 represent 33.8 and 13.4% of the dataset variance respectively. 95% confidence ellipses are shown.

List of Tables

Abstract

Lysosomal storage disorders (LSDs) are predominantly rare, autosomal recessive, orphan diseases that impact the storage of specific molecules within the lysosome. This work explores current metabolomics research on LSDs with a specific focus on Niemann Pick Disease Type C, Gaucher's Disease and Gangliosidoses. Moreover, metabolomic techniques are explored with a history and theory outlined for nuclear magnetic resonance (NMR) and liquid chromatography-mass spectrometric analytical strategies. Statistical analyses, particularly multivariate approaches, and machine learning applied to the analysis of these metabolomic datasets are also discussed. Human urinary and plasma metabolomics investigations were then performed using the ^{1}H NMR technique on samples provided from Gaucher's Disease and GM1 Type II patients with corresponding controls. Furthermore, using a feline model, a ^{1}H NMR urinary investigation was performed to investigate the impact of Niemann Pick Disease Type C on the metabolome, and also to evaluate metabolic changes in felines treated with 2-hydroxypropyl-β-cyclodextrin. Statistical evaluations were performed for all three conditions, and the significance of these are highlighted, with some key potential biomarkers identified. The advantages of using a ^{1}H NMR approach are summarised, in addition to the challenges of performing complex metabolomics analysis of LSDs, with particular considerations of the limited sample sizes of these available for such studies.

Chapter 1: Literature Review

1.0 Introduction

In the functional lysosome, lipids and other macromolecules, such as carbohydrates, peptides, nucleic acids *etc.*, are degraded via catabolism through the action of acid hydrolases. However, inherited defects in genes that encode lysosomal enzymes and non-enzymatic proteins can cause the build-up of macromolecules within the late endocytic system [2]. The 'umbrella' term for these diseases is lysosomal storage disorders (LSDs), and these can be further sub-categorised into lipid storage disorders (LLSDs or lipidoses), mucopolysaccharidoses, glycoprotein storage disorders and mucolipidoses. LLSDs are the focus of this review, and are listed in Table 1.0, together with their corresponding defective genes encoding the specific enzyme/protein which is depleted or malfunctional, and any diagnostic tests available. Niemann-Pick type C (NPC) disease is caused by mutations in two genes *NPC1* (95% of clinical cases) and *NPC2* and exhibits a different pathological expression to those of other LLSDs, in that the transportation of lipids and cholesterol is predominantly affected. Indeed, a complex storage pattern of multiple lipids within late endosomes/lysosomes represents a characteristic feature of this condition. It appears that the lysosomal storage of sphingosine gives rise to a loss of lysosomal calcium ions, leading to the secondary accumulation of unesterified cholesterol and glycosphingolipids. *NPC1* encodes a large membrane glycoprotein that resides in the limiting membrane of the lysosome and is termed *NPC1*. *NPC2* encodes a small soluble lysosomal protein that binds cholesterol within the lysosomal lumen, and which is believed to transfer cholesterol to the NPC1 protein.

Lipid Storage Disorder	**Defected Gene**	**Diagnosis**	**Approved Therapies**
GM1 Gangliosidosis	*GLB1*	- Patient history - Clinical features - Symptomatic approach - Confirmation of increased storage or genetic tests - Typical approaches to confirm storage are enzyme-based	S/S
Sandhoff Disease (GM2)	*HEXB*		S/S
Tay-Sachs Disease (GM2)	*HEXA*		S/S
Niemann-Pick type A & B (NPA & NPB)	*SMPD1*		S/S
Niemann-Pick type C1 (NPC1)	*NPC1* (95%)		SRT
Niemann-Pick type C2 (NPC2)	*NPC2* (5%)		SRT
Gaucher's Disease (GD) types I-III	*GBA*		ERT and SRT

Fabry Disease (FD)	*GLA*	- Some tests are invasive and require a skin biopsy	ERT and CT
Krabbe Disease (KD)	*GLB*		HSCT and S/S

Table 1.0 Lipidoses (LLSDs) with corresponding defective genes (adapted from [3]). Abbreviations: CT, Chaperone Therapy; ERT, Enzyme Replacement Therapy; HSCT, Haematopoietic Stem Cell Therapy; S/S, Symptomatic and Supportive Therapy; SRT, Substrate Reduction Therapy.

For these debilitating disorders, mortality incidence is often high at a young age, i.e. within a few months or years of birth. However, selected disease-specific variations may allow sufferers of some of these conditions to live into adulthood. For example, NPA is the most severe of the three types of Niemann-Pick (NP) conditions, whereas in NPC disease, patients may display no signs and symptoms until adulthood. LLSDs display a wide range of symptoms, and this renders diagnosis a challenging process. Furthermore, LLSDs range in severity and onset, but overall life expectancy for them is low. No single symptom can robustly diagnose LLSDs; instead, a combination of patient history, observed and monitored symptoms, and more specific testing is required in order to obtain a reliable diagnosis [4], [5] as noted in Table 1.0. Often, untargeted metabolomics analysis strategies are able to provide an insight into early markers for diseases, and hence valuable contributions provided by the applications of such techniques to LLSDs may be vital, since some diagnoses can be onerous and take a considerable period of time.

At present, the applications of multivariate (MV) metabolomics strategies have been comprehensively reviewed for many diseases and their biomarkers [6], for example specific reviews focused on cancer [7], [8] and Parkinson's Disease [9]. The core linkages between LSDs and other neurological disorders have also been described [2], [3]. Interestingly, the chemopathological aetiologies and developments associated with GD and gangliosidoses may give rise to an improved understanding of Parkinson's disease, inclusive of neuropathological, genetic and clinical features within disease groups [2], [3]. An increase in blood serum gangliosides has also been related to cancer, and GD3 synthase has been suggested as a novel drug target for cancer [10]. Hence, an improved understanding of LSD processes has the potential to enhance our knowledge of both related and unrelated disorders too.

This literature review will consider how metabolomics strategies have been applied to LLSDs, and the relative merits of these applications, including how such approaches have developed our current understanding of these disorders, and their role as a technique in the development of new and innovative therapeutic strategies. Our focus will include high-resolution nuclear magnetic resonance (NMR) spectroscopic, mass-spectrometric (MS), and

respective hyphenated techniques *i.e.* liquid chromatographic-mass spectrometric (LC-MS) and LC-MS/MS analyses. Applications of metabolomics techniques to LSD biomarker research has increased since the late 2000s, and full details of all contributions made are documented in Chapter 2. These methodologies for candidate biomarkers are not only assisting in or confirming disease diagnosis, but also their severities and scope. Often, elevations or depletions of one or more biomarkers is symptomatic of disease severity [11], and therefore the monitoring and prognostic stratification of these diseases may be possible. Currently, a range of biofluids and biomarkers have been considered in both human and animal trials, and these will be discussed in depth for GD, NPC and gangliosidoses.

1.1 Niemann Pick Diseases

Most of the metabolomics research conducted on sphingolipid storage diseases has been focused on NPC1 disease [12]. High-resolution ^{1}H NMR analysis has been successful in identifying potential urinary biomarkers for NPC1 disease diagnosis, and these include branched-chain amino acids (BCAAs), selected bile acids (BAs) and 3-aminoisobutyrate [13], the latter arising from either BCAA catabolism or thymine degradation. This analysis further indicated that the brain and liver were the prominent sites affected, which is consistent with the pathology described above. with area under the curve (AUROC) values ranging from 0.81-0.91 for the most significant metabolites, and partial least squares-discriminatory analysis (PLS-DA) yielding significant values of Q^2 (0.56) and an accuracy of 0.93 [13]. This approach has potential applications for disease progression monitoring, but these benefits have yet to be demonstrated in practice. At the sub-cellular level, the lysosome, ER and mitochondria were the main organelles affected in this investigation. Moreover, high-performance liquid chromatographic (HPLC) analysis coupled with a UV detection system has also been utilised to target specific amino acid imbalances and DNA methylation in murine cerebellum collected from experimental mice with this condition, in which BCAAs significantly ($p > 0.05$) arose from elevated glutamine/glutamate oxidation, in a response to prolonged neuron survival [14]. The combination of DNA methylation, which corroborated amino acid imbalances, explores new mechanisms which are ascribable to neurodegeneration in NPC disease. Indeed, this combinatorial approach could influence new treatment strategies, in which these amino acid imbalances are addressed by specific treatments.

LC-MS analysis has been used to understand upregulations or downregulations of potential biomarkers, as well as changes in biomarker expression in NPC1 disease as a response to treatment with miglustat and 2-hydroxypropyl-β-cyclodextrin (HPβCD) [11]. This study

monitored the effectiveness of drug administration and uptake, and identified specific upregulated markers in mouse brain, liver and spleen. These included elevated levels of ceramide, GM1, GM3, sphingolipid species, and lactosylceramide in NPC1 tissues, which markedly reduced upon treatment with HPβCD, comparatively to miglustat [11]. Elevated markers in human plasma and CSF were identified as monohexosylceramides and ceramides [11]. It was demonstrated that upon administration of the substrate reduction therapy drug miglustat (whereby inhibition of GlcCer synthase occurs) these markers did not significantly reduce [9]. However, it was proposed these biomarkers may downregulate on administration of HPβCD, as this was observed in the tissues of the mouse model [11]. This research needs further confirmation of biomarkers prior to translation into clinical practice for disease diagnosis or clinical progression. Another research study, [15], also monitored the effectiveness of treatment with HPβCD in mice, using 2,4(S)-hydroxycholesterol (2,4(S)-HC) as a pharmacodynamic marker in a National Institute of Health (NIH) clinical trial, and employed the LC-MS/MS technique to assess metabolic activity in NPC1 disease. Brain 2,4(S)-HC build-up was found to arise from high cholesterol levels. 2,4(S)-HC is produced exclusively at this location, and acts as a biological response to eliminate cholesterol; excess cholesterol is toxic, and 2,4(S)-HC can cross the blood-brain barrier (BBB), unlike cholesterol itself [15]. A different investigation, [16], also monitored 2,4(S)-HC levels in blood plasma in order to assess treatment outcomes with HPβCD in murine models, and from these studies proposed this agent as a biomarker for diagnostics and therapeutics development.

Oxidation products of cholesterol are commonly monitored biomarkers in NPC disease, not only because of the increased levels of intercellular free cholesterol, but also in view of oxidative stress which is a feature of the condition. Similarly, selected BAs also serve as disease biomarkers; since these are also synthesised from cholesterol in the liver, they have been correlated to cholestasis in NPC1 patients [13], [17]. One study, [17], has reported increases in of cholestane-3β5α6β-triol levels in NPC1 patient plasma profiles, and in mouse plasma and liver, whole brain and cerebellum tissues. Cholestane-3β5α6β-triol, a further oxidation product of cholesterol, which functions as an endogenous neuroprotectant, has been reported to be upregulated in the blood plasma of NPC patients [18]–[21]. Other disorders, such as lysosomal acid lipase deficiency and cerebrotendinous xanthomatosis, have also been shown to correlate with enhanced cholestane-3β5α6β-triol concentrations in human plasma, and hence this biomarker suffers specificity issues for the diagnosis of NPC1 disease [20]. Others indicated that cholestane-3β5α6β-triol could be used as a biomarker in conjunction with lyso-sphingomyelin-509, an approach which provided a model with 91% specificity and 100% sensitivity [21]. These findings were validated using HPβCD, with

cholestane-3β5α6β-triol significantly downregulated with treatment [17]. However, such critical effective drug-based validation processes were not performed in the studies of [18]–[21].

Two investigations, [17], [18], also reported an elevation in 7-ketocholesterol, which is also an oxidation product of cholesterol, indicating that this may be non-specific as a biomarker for diagnosis, as suggested in other studies [20], [22]. Human plasma 7-ketocholesterol concentrations were also found to be elevated in NPB, sterol disorders, peroxisomal diseases, metachromatic leukodystrophy, mucolipidosis II/III, LAL-deficiencies [20] and cardiovascular diseases [23]. Notably, other correlations between NPC and NPB patients have been noted, since both exhibit elevated levels of lysosphingomyelin in blood plasma and dried blood spots as a diagnostic marker [24], [25]. Indeed, current literature predominately focuses on biomarkers known to be associated directly with the disorder, and the accumulation of intracellular metabolites, rather than what effect this has on the metabolome as a whole. A combination of statistically significant biomarkers, such as cholesterol oxidation products, BAs and 3-AIB, inclusive of those which are directly and indirectly associated with the disease, could offer a more sensitive and specific solution for the long-term monitoring of LSDs [26]. Other potential sphingomyelin precursors have been investigated including sphingosylphosphorylcholine, which has been demonstrated to be marginally increased in NPC plasma using an LC-MS/MS strategy in a pilot study [27]. However, it was recommended that this biomarker should be used in conjunction with for example plasma lysosphingomyelin-509, also known as N-palmitoyl-O-phosphocholineserine [28] (with a median fold-change value of 65 in NPC plasma relative to healthy controls), in order to be applied in practice as a significant biomarker for diagnostic purposes [27].

Further studies have been performed using LC-MS-based methodologies, analysing no more than 2 NPC samples which monitored urinary bile acid sulphate conjugates for diagnostics, i.e. 3β-sulfooxy-7β-hydroxy-5-cholen-24-oic acid and 3β-sulfooxy-7-oxo-5-cholen-24-oic acid, both of which were found to be very significantly upregulated in NPC disease [29]–[31]. However, more recently, a larger study was performed by the same group and demonstrated upregulations in 3β-sulfoxy-7β-hydroxy-5-cholen-24-oic acid and its glycine and taurine conjugate, using urine samples from NPC patients (n = 23) and control (n = 28) participants [32]. A further investigation with control (n = 38) and NPC patients (n = 28) found similar findings with an AUROC value > 0.92 for 3β-sulfoxy-7β-hydroxy-5-cholen-24-oic acid, 3β-sulfoxy-7-oxo-5-cholen-24-oic acid, non-amidated 3β-sulfooxy-7β-N-acetylglucosaminyl-5-cholen-24-oic acid, glycine-amidated 3β-sulfoxy-7β-N-

acetylglucosaminyl-5-cholen-24-oic acid, and taurine-amidated 3β-sulfoxy-7β-N-acetylglucosaminyl-5-cholen-24-oic acid [33]. The most significant marker with an AUROC value of 1.0 was 3β-sulfooxy-7β-hydroxy-5-cholen-24-oic acid [33]. The sensitivity and specificity of these biomarkers, and how treatments can affect them, needs to be explored.

BAs have been identified as potential biomarkers for early diagnosis in view of the observed upregulation of their precursors 7-oxocholesterol and cholestane-3β5α6β-triol in human urine and plasma. A novel MS assay was created in order to detect these bile acid biomarkers in dried blood spots, urine and plasma, and with a much larger sample size than those featured in the above studies [34]. Elevated BA concentrations indicate liver dysfunction; however, there not as yet definitive evidence for a decrease in these biomarkers following disease-modifying treatments, and hence they require further validation [34].

A combination of these biomarkers in selected biofluids could provide more statistical power, since the majority of the analyses performed in these studies were targeted and univariate, whilst MV approaches could provide improved specificity for NP disease diagnostics. However, few investigations have evidenced these markers against disease progression and disease severity indices, which are clearly of much importance for clinical management in this disorder [35]. It is therefore necessary to ensure that potential biomarkers selected for NPC1 disease are specific to this condition in order to avoid misdiagnoses. At present, there is a lack of evidence in the literature to show the intrinsic value of these candidate biomarkers individually. Indeed, most of these diseases exhibit subtly different manifestations, translating into modified metabolic responses. Therefore, without thorough testing of the range of other LSDs, and assurance of the validity of such biomarkers, these cannot be entirely confirmatory, and hence misdiagnosis remains a possibility.

Studies performed on Chinese hamster ovary cells, and those collected from an NPC patient versus an NPC carrier, and also COS-7 and fibroblast cells from an NPC patient, have also been analysed using an HPLC electrospray ionisation (ESI)/MS/MS strategy to determine sphingomyelin levels [36]. This study demonstrated an upregulation in total sphingomyelin, including specific species, such as d18:1/18:1, d18:1/18:0, d18:1/20:0, d18:1/22:0, d18:1/24:1, and d18:1/26:1 [36]. Moreover, this investigation showed that Rab9, a GTP-binding protein that controls trafficking from the late endosome to the Golgi, is defective [36] and upon reducing sphingomyelin within the cell using sphingomyelinase, Rab9 was able to function more effectively, and cholesterol accumulation was concomitantly diminished significantly [36]. Consequently, sphingomyelin reduction is another potential therapeutic

target for NPC disease. This demonstrates the usefulness of metabolomics analysis of cell systems in lysosomal storage disorders to assess the mechanisms of cellular pathogenesis.

Finally, high-resolution NMR analysis has been employed to observe the urinary excretion of miglustat, valproate, and valproate metabolites in NPC1 disease [37], which could contribute to more personalised medicine approaches in human patients. Similar studies have also been performed with high-performance liquid chromatography coupled with a light-scattering detection system (HPLC-LS), and these demonstrated the detection of HPβCD in human urine [38].

1.2 Gaucher's Disease

Similar to NPC disease, there are variants of Gaucher disease (GD) (types 1, 2 and 3): GD1 (*OMIM 230800*) is non-neurological, whilst GD2 (*OMIM 230900*) and GD3 (*OMIM 231000*) are neurological conditions also involving the spleen and liver. Targeted metabolomics approaches have also been applied to GD, using LC and LC-MS technologies. The plasma biomarker glucosylceramide (GlcCer), and ceramide ratios, have been used in combination with ERT treatment for this condition, with levels significantly decreasing with corresponding treatments in GD1 patients [39]. Moreover, glucosylsphingosine has also been determined to be a biomarker in dried blood spots (DBS) from children with GD1 and 3 diseases [48]. Glucosylsphingosine has also proven to have diagnostic potential in in other LSDs, further validating its potential as a generalised LSD-specific biomarker [40], [41]. However, as discussed by [24], glucosylsphingosine is also markedly elevated in the plasma of patients with NPC disease, but to a different extent - GD concentrations of glucosylsphingosine are relatively higher [24]. Another investigation showed an upregulation of glucosylsphingosine in Krabbe Disease (KD) and GD plasma samples, but the concentrations were distinguishable, since GD patients had significantly higher levels [42]. If this were the biomarker of choice, essential reference ranges would ensure that the correct diagnosis was given. Urine has received little attention as a potential biofluid sample type for GD metabolomics studies.

Another investigation, [43], reported a potential marker for GD in serum, specifically GlcCer. Notwithstanding, there were validation issues in view of the lack of controls and heterozygous carrier samples integrated into the study for comparative purposes. Indeed, with a sample size of 2, the results presented by [43] also raises concerns regarding the reliability of results acquired. However, this represents one of the challenges presented by

the monitoring a rare LSD in human participants. Furthermore, samples were collected using invasive methodologies. Invasive sample collection is a common theme in the observation and monitoring of LSDs; more data could be generated by considering urinary metabolic profiles, together with those of other readily-accessible biofluids such as saliva, perhaps, which has successfully been utilised for the diagnosis of other diseases, including cancer and diabetes [44].

1.3 GM1, GM2 and GM3 Gangliosidoses

Although much is known regarding the clinical aspects of the GM1 *(OMIM 230500)* (GM1), and GM2 gangliosidoses (*OMIM 268800*) (GM2), the cellular mechanisms leading to neurodegeneration are poorly understood, and the precise functions of gangliosides (i.e., their cellular and biochemical properties) remain incomplete. Moreover, very few metabolomics studies are yet to be performed on the GM1 and GM2 gangliosidoses.

More recently, reversed-phase LC has been utilised to investigate potential biomarkers that can be analysed in biofluids by observing those detectable and validated in brain and liver tissues from mice, and hippocampus samples from humans with Sandhoff (SD, GM2) disease [45]. PCA analysis was able to successfully distinguish between SD and control hippocampus samples from humans, in addition to brain and liver samples from control and SD mice [45]. Disturbances were observed in pathways associated with protein catabolism and lipid metabolism. N-acetyl compounds were altered in the SD brain from humans and mice, including elevated N-acetylgalactosamine 4-sulphate which has been proposed to be correlated with a deficiency in the HexB enzyme [45]. N-Acetyl-L-aspartate and L-glutamate were downregulated in the brain [45]. The investigation focussed on brain samples and provided potential new targets for biomarkers which may be accessible in non-invasively collected biofluids.

A study [46] distinguished between different gangliosides in a study which unfortunately did not account for the matrix effects exerted by biofluids, since the experiments were performed *in vitro*. Another investigation, [47], went further and developed a liquid chromatography-electrospray ionisation-mass spectrometry (LC-ESI-MS) methodology which was able to monitor blood serum ganglioside levels in participants (n = 10) with different oesophageal diseases, but not the gangliosidoses. Other authors, [48], have developed a methodology for the observation of GM1 and GM2 gangliosides simultaneously in CSF using LC-MS-MS. This methodology used 30 human CSF samples in which GM1 and GM2 were 'spiked' and

recovered with a high accuracy, i.e. 98 and 102% respectively. More recently, [49], employed quantitative immunoassays to determine key biomarkers in CSF and blood serum in 2 infantile Sandhoff, 9 infantile Tay-Sach's, 5 juvenile Sandhoff, 6 infantile GM1, 7 late infantile GM1 and 1 juvenile GM1 patients, and compared these to 100 healthy control participants. Of 188 metabolites tested, five key markers identified were epithelial-derived neutrophil-activating protein 78, monocyte chemotactic protein 1, macrophage inflammatory protein-1 alpha, macrophage inflammatory protein-1β, and tumour necrosis factor receptor 2, which were all upregulated in the CSF of gangliosidosis infantiles, and were also linked to inflammation within the central nervous system (CNS). In view of their rarity and severity, longitudinal metabolomics studies will, of course, take longer to perform. Markers which could be evaluated include those pertaining to CNS degradation, or lysosomal dysfunction such as quinolinate and N-acetylsugars respectively.

Another investigation, [50], explored urinary N-acetylated storage compound biomarkers in patients with GM1 disease which arises from β-D-galactosidase deficiency. These researchers employed preparative TLC sequentially followed by 1D ^{1}H NMR analysis, and found that the first of two bands investigated in this manner corresponded to an N-acetylated biantennary octasaccharide, and the second to an N-acetylated triantennary decasaccharide or tetraantennary dodecasaccharide [50]. The 500 MHz ^{1}H NMR profile of a patient with this condition contained a series of N-acetyl-CH_3 function singlet resonances within the 2.02-2.04 ppm range at an analytical pH value of 2.50, that with the highest intensity being located at δ = 2.03 ppm. These results are consistent with elevations in the urinary concentrations of oligosaccharides with a D-galactosyl residue at their non-reducing ends.

^{1}H NMR spectra acquired on urine collected from a patient with Sandhoff disease demonstrated the excretion of oligosaccharides, specifically acetamido-C$\underline{H}_3$ function singlet resonances within the δ = 2.02-2.08 ppm range, the highest intensity one corresponding to a concentration of *ca.* 80 μmol/mmol of creatinine. Combined preparative TLC-^{1}H NMR analysis provided evidence that these signals were at least partially ascribable to oligosaccharides with the structures β-GlcNAc-(1→2)-α-Man-(1-3)-[β-GlcNAc-(1→2)-α-Man-(1-6)]-β-Man-(1→4)-GlcNAc (band 1), and β-GlcNAc-(1→2)-α-Man-(1-3)-[β-GlcNAc-(1→4)][β-GlcNAc-(1→2)]-α-Man-(1-6)]-β-Man-(1→4)-GlcNAc (band 2).

1.4 Future Perspectives for and Potential Limitations of Metabolomic Investigations for Monitoring LSDs

Although [51] recently discussed some future perspectives for metabolomics in inborn errors of metabolism, these considerations could be further developed. For example, a thorough metabolomics analysis of the effects of ERTs and SRTs on patients with LLSDs could provide vital insights into the effectiveness of these clinical approaches by observing biomarkers in a multidimensional context, a simple example being the concomitant observation of elevated markers decreasing, and depleted markers rising in their biofluid concentrations. It is evident that new, robust, non-invasive treatment options need to be explored for LLSDs: in the case of neurological disorders, variants of these therapies are required to successfully cross the BBB. To enhance ERT, the use of nanoparticles to facilitate the ability of enzymes cross the BBB is also being explored, since they protect the enzyme delivered against its in vivo degradation [52]. Thus, metabolomics could play a role in ensuring that this methodology is effective, and perhaps also serve to validate candidate biomarkers. Moreover, nanoparticles are also being employed for the gene therapy trialling of non-viral vectors, and again metabolomics could facilitate the provision of insightful information regarding the effectiveness of such therapies [52]. However, one major potential drawback for the use of nanoparticles is that their long-term toxicity remains speculative, a process which again could be explored using metabolomics as a vital probing tool.

Urine samples are yet to be analysed to the same extent as blood plasma/serum and CSF in these patients, and donations of this biofluid obviously constitutes a less invasive collection procedure. Although all biofluids and tissues have been probed in these works, plasma appears to be the biofluid of choice for such studies. Interestingly, no metabolomics studies have been performed on human saliva, which would be advantageous for virtually non-invasive sampling purposes. In principle, saliva also has the potential to represent the full physiological status of the human body, just as blood serum or plasma samples can, it should be noted that many fully-characterised, established blood plasma markers are often present in saliva at much lower concentrations [44]. Salivary biomarkers have been found using metabolomics in diseases, such as cancer and diabetes [44]. However, the disadvantage of use of this biofluid for such investigations is that it may easily be compromised, i.e. influenced by external factors such as poor oral health/hygiene, and eating or drinking episodes prior to sampling. Hence, these investigations should be carried out with extreme care, particularly careful sample collection [53]. Moreover, very few studies have been performed on brain tissue from living participants in view of the inability to retrieve such samples. However, murine brain has been analysed as described above [15], [17], [45]. However, currently there remains a whole scope of unexplored biofluids, tissues and cells that can be investigated for LSDs, e.g. urine and CSF in Gangliosidosis. NP has been investigated by MS- and NMR-

linked metabolomics investigations thoroughly, with a total of 24 publications arising from the study of the disease. Less metabolomics research has been performed on GD, however, with only 5 publications targeting the disease, respectively at the time of this report, and very little has been completed on Gangliosidoses. These are areas which could clearly be further investigated. It would also be of much interest to explore metabolic correlations of both biomarker and non-biomarker variables between differing sub-classes of the same disorder, i.e. NPA, NPB and NPC, and any similarities they may have with GD and gangliosidosis. This has shown to be extremely useful in other investigations to establish the specificity of the marker [54].

Lack of uniform studies across all LLSDs is evident within this review in terms of experimental design and statistical approach, not to mention the bioanalytical strategies and techniques employed for these purposes. More systematic metabolomics studies will provide a wealth of information regarding the significance of the current biomarkers identified, and also allow for more facile and expedient cross-validation of metabolites. Larger scale studies are also required, since they aim to provide much more representative dataset profiles. [34]'s study is a good example of this. As discussed above, it is also of much importance to use a biomarker or combination of biomarkers which are rigorous enough to avoid misdiagnosis, as demonstrated by [20], [22].

LC-MS/MS is clearly showing some promise in detecting NPC disease, demonstrating an elevated level of blood plasma cholestane-3β5α6β-triol. However, there is a major requirement for such developments to be integrated into clinical practice, perhaps in combination with one or more specific biomarker for potentially confounding sterol disorders, or for NPC disease itself. A combination of the phenome, metabolome and genome could translate the research in metabolomics into practice to modernise medicine; however, for this to operate effectively, method development and validation are also crucial [6]. One of the major disadvantages of translating MV NMR datasets into medical practices is that the facilities employed for these purposes, such as NMR and LC-MS, have both high purchase and maintenance costs, and require specialist training for use, interpretation and subsequent data analysis. However, NMR technology is also showing some promise in benchtop instrumentation, which could be considered for clinical applications with the advent and development of water suppression sequences and automation now possible [55]. NMR approaches may prove to be invaluable as substantial technological advancements are being made in this field, with technologies such as hyperpolarisation increasing instrumental sensitivities. Emerging technology such as computational tools are also becoming a more integral part of metabolomics applications, and artificial intelligence technology could

effectively simplify and distinguish normal from unusual patterns of results acquired by metabolomics-based instrumentation.

Another area of metabolomics research which is yet to be explored in depth is neonatal and prenatal testing for LLSDs. Some investigations have been performed on paediatric patients, and to date this offers some opportunities for further developments within this field [56], [57]. These diseases are often overlooked in view of their rarity, and hence diagnosis and prognosis are not established until later in the lifespan of the patient. Notwithstanding, currently available reports have emphasised that early detection of these diseases is a highly beneficial patient requirement, and hence early testing and diagnosis could indeed represent a life-saving process.

In summary, despite some progress, metabolomics approaches have yet to influence diagnosis and monitoring of LLSDs in clinical practice. Clinical diagnosis is dominated by PCR-based genetic testing and enzyme assays, and hence using metabolomics to diagnose LLSDs is still in its infancy. Nevertheless, future opportunities in this area remain, most especially in an era where personalised medicine is becoming an increasingly greater practical likelihood. Future developments required for LLSDs include (1) systematic sample collection and handling, (2) further validation of at least some of the potential biomarkers suggested via the rigorous monitoring of their biofluid levels along with simultaneous co-ordinated assessments of clinical responses to treatment(s); (3) detailed correlations of biomarker concentrations with disease severity, and also demographic variables such as age, gender and diet, etc.; and (4) a further detailed review in order to ensure that biomarker(s) identified are solely specific to particular LLSDs.

1.5 Conclusions

To date, metabolomics techniques have only scratched the surface in terms of their applications to the lipidoses; however, to date they have provided some promising diagnostic and prognostic information for NPC-1 disease and its responses to different therapeutic treatments, which has revealed some promising biomarkers. In addition, more global experimental design and bioanalysis strategies may provide more valuable information, rather than limiting research to that focused only on a limited number of biomarkers as in pre-targeted methodologies. Moreover, carefully-planned meta-analysis studies involving MV adjustments may also offer support for both targeted and untargeted LSD biomarker research. This thesis aims to explore these developments further.

Chapter 2: Methodology

2.1 Metabolomics

There are over 200 'omics' sciences in existence, in which metabolomics is a relatively new field of these and this involves the characterisation of small molecular 'signatures' or 'fingerprints' which are associated with biological systems [58]. The metabolome itself consists of low-molecular-mass compounds which are present in certain cells at a particular phase in the organism studied [59]. All molecules are involved in metabolic processes once inside an organism, and metabolism is a set of reactions which cells perform for growth, maintenance, reproduction and death. Disruption of these processes can cause problems within the system analysed.

Metabolomics plays a key role in combination with genomics and proteomics, as illustrated in Figure 2.0 and promotes our understanding of cellular processes, diseases, and mutations, but the metabolomics approach is quite different, since these older 'omic's techniques are expensive, labour-intensive, and cannot easily contend with the thousands of biomolecules which are simultaneously identified, quantified and statistically differentiated [59]–[62]. Unlike genomics and proteomics, metabolomics focuses on biological fluids which express the pathological status of that organism at a particular point in time, which can change in a diurnal or more prolonged time-dependent manner in view of environmental factors, and therefore metabolomics investigations are very complex [59]. However, metabolomics strategies are very advantageous, providing detailed information regarding cellular functions within the organism, and also the effects of xenobiotic exposure [62].

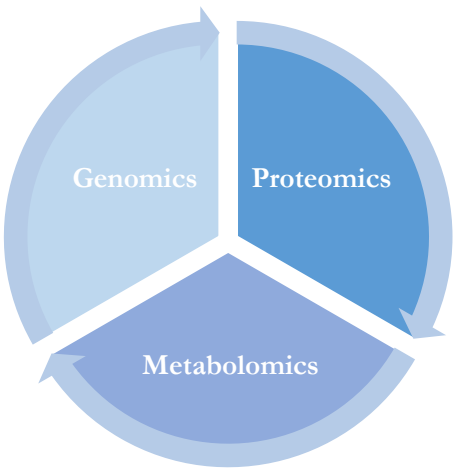


Figure 2.0. Loop diagram representative of the relationships between genomics, proteomics and metabolomics.

Metabolomics studies are vital in healthcare since they aim to observe disease-specific metabolic modifications, and in in particular highlighting specific biomarkers in a metabolic profile relating to disease, and enhancing disease diagnostics *via* providing knowledge regarding which biological pathways are being expressed [6], [63]. In addition, metabolomics analysis has potential to show the effectivity of the treatments prescribed alongside their downstream products, which can aid pharmaceutical research and development [63]. Metabolomics has alternative applications in nutrition, environmental studies, agriculture and food safety, but these are not explored in this particular research work [63], [64].

The growth of this field over the past 20 years is astonishing and ever-expanding, with a focus on an entire biological system rather than just a single cellular system [62]. Metabolomics was first a relatively common practice in ancient Chinese medicine which noted different tastes, smells and colours of urine [64]. This biofluid, which is considered a waste product from the body, was further recognised as extremely useful in diagnosis of disease by the Egyptians, who noted that the smell, taste and colour of urine was different with different medical conditions and could be 'mapped' (Figure 2.1) [60]. Urine is composed of 95% (w/w) water, but also contains ions such as sodium, chloride, potassium, phosphate, sulphate, iodide, amides and also a number of amino acids and organic acids present at low concentrations [65]. Colour wheels were soon developed, which gave diagnostic information indicative of diseases; for example, a brownish colour would indicate jaundice [60].

Although the ancient Chinese and Egyptians utilised these methodologies originally, metabolomics was first coined in 1940 by Roger Williams, and by using chromatography, he found a difference between the urinary profiles of schizophrenics when compared to those of healthy individuals [66]. Much later, Oliver in 1998 noticed differences in yeast cultures

with different genomes using microbiology, and termed this as metabolomics, a descriptor which was refined by Nicholson in 1999 [67].

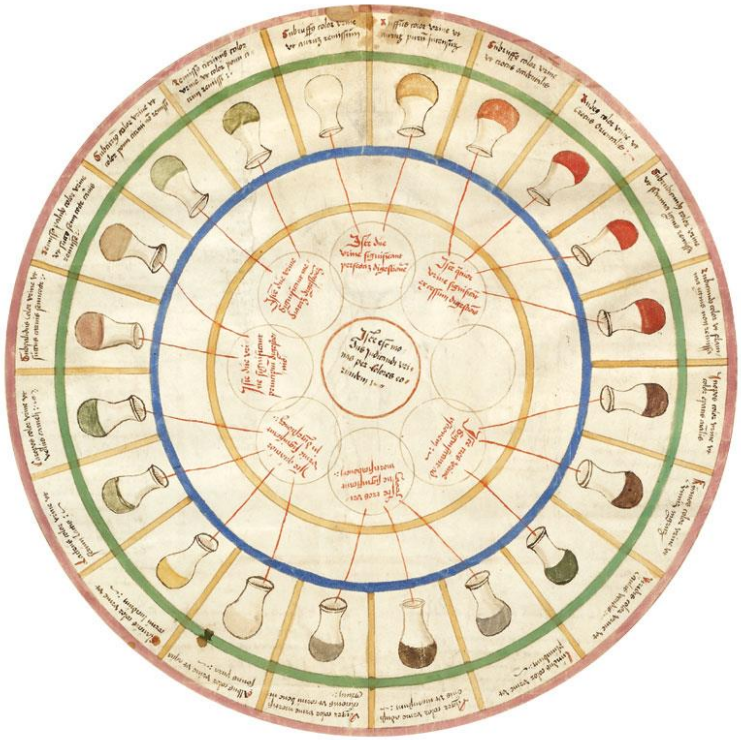


Figure 2.1. Egyptian urine wheel describing colours, tastes and smells of urine, published in 1506 by Ullrich Pinder in his book *Epiphanie Medicorum.*

Urine gained even more diagnostic recognition when it was then used for diagnosis of the first genetic disease - alkaptonuria [60]. Nowadays, urine is collected in universal containers and more routinely used for dipstick tests which can instantaneously measure urinary glucose, bilirubin, ketone bodies, nitrates, leukocytes, haemoglobin etc. [60], (Figure 2.2) with more detailed urinalysis being utilised for renal conditions. Furthermore, it is frequently used as a metabolomics tool [60]. For example, increased urinary levels of taurine and creatine have been correlated with acute liver and testicular toxicity respectively, but metabolomics is not limited to just these examples, since there are many others [62]. The primary advantage of urine as a sample is that it is considered non-invasive for diagnostic purposes, has a simple collection procedure, and large volumes of it are readily available.

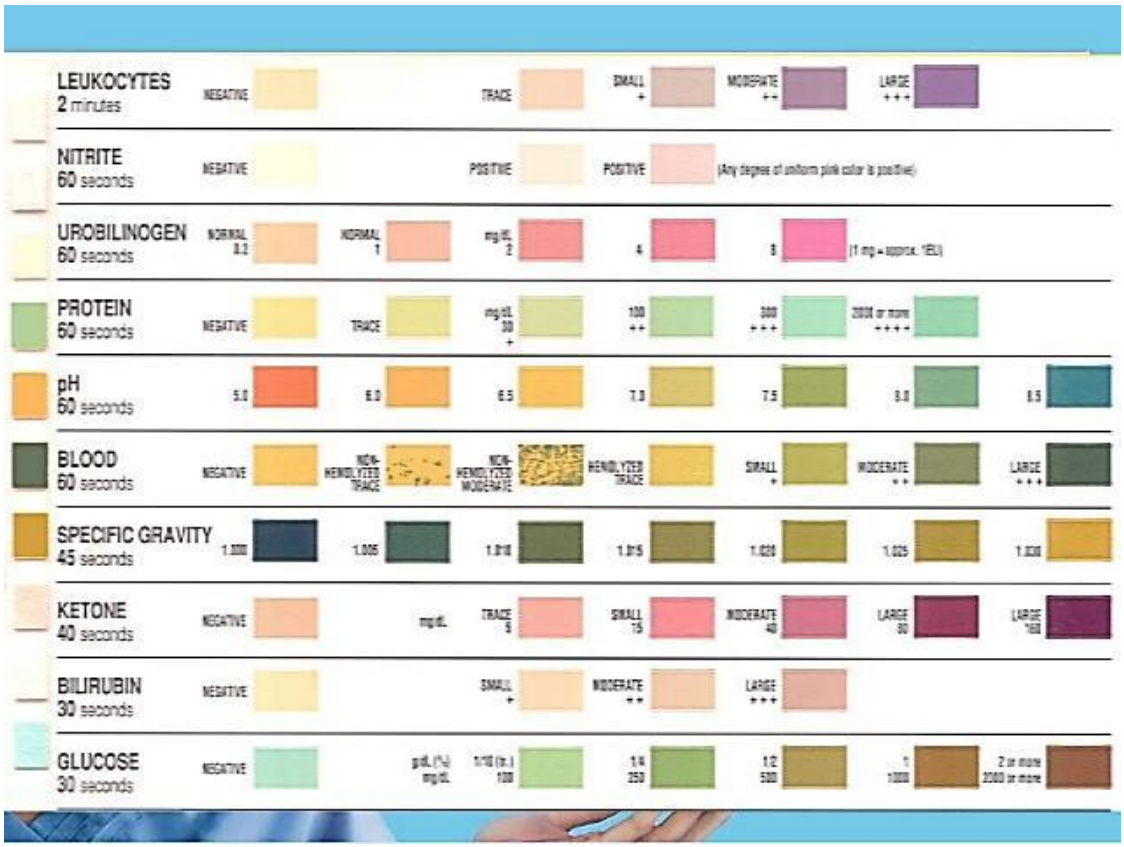


Figure 2.2. Typical dipstick test information derivable from human urine samples.

Indeed, although urine demonstrated diagnostic application potential, blood has also shown to be extremely valuable for these studies. This biofluid, which is composed of plasma/serum and cellular matter, is predominantly water-containing, and has a primary function of transportation of cells and molecules including platelets, white and red blood cells, and proteins around the body [61]. Although blood is useful in metabolomics, it is more difficult to collect than urine, since it is collected by venepuncture by a trained healthcare specialist (phlebotomist).

Both plasma and serum can be utilised for metabolomics analysis. Blood can be collected in combination with an anticoagulant, for example lithium heparin tubes or ethylenediaminetetraacetic acid (EDTA) tubes, and after centrifugation, plasma is collected. Without the anticoagulant, serum can be collected [61]. The primary responsibility of blood is to behave like the body's 'motorway', transporting hormones, gases, nutrients and metabolic waste to appropriate organs [61].

Blood has been analysed for over 70 years, and has profound clinical importance; indeed, reference ranges have been established for the majority of its components, and tests similar to those shown in Figure 2.2 have been developed for blood, and metabolomics analysis is

growing for both serum and plasma [61]. Some diseases are a direct influence of the condition of the blood such as cancers such as leukaemia, myeloma and lymphoma.

Cerebrospinal fluid (CSF) studies are more infant than those involving blood and urine, although over the past 60 years reference ranges and metabolomics studies have been applied to CSF [63]. CSF is less commonly used as a sample for metabolomics purposes, since it has a more invasive sample collection nature, which is performed via lumber puncture [68]. CSF is located in the ventricles and subarachnoid space of the brain and spinal column, and has the primary function of protecting the brain from shock, but it also manages the nutrients and biomolecules filtered from the blood to the brain [63]. In view of its location, CSF is commonly utilised for diseases which affect the central nervous system (CNS), and plays an important role in biomedical research and clinical chemistry for diagnosing and monitoring diseases such as Parkinson's and Alzheimer's diseases, multiple sclerosis and brain injuries [63], [69]. LLSDs also affect the CNS, and CSF is yet to be analysed in metabolomics studies exploring such conditions as described in Chapter 1.

At present, all biofluids, cell extracts, tissues and even organs are used in metabolomics research, and provide researchers with key information regarding the pathological status of patients, information which other 'omics' fields fail to fulfil [64]. More recently in 2004 the Human Metabolome project was launched in order for scientists to gain further understanding regarding metabolites in the human body, and in order to harmonise information gained across different bioanalytical platforms [65]. Databases have been developed to increase the ease of sharing information when new metabolites are found, and these include *Biological Magnetic Resonance Bank* (BMRB) and *The Human Metabolome Database* (HMDB), which integrates The Urine Metabolome Database (UMD) and The Serum Metabolome Database (SMDB), has the ability to enable scientists to identify thousands of compounds in a range of biofluids, this number continually growing as the sensitivity of techniques available increases, together with the number of diseases being analysed via this process [60], [65].

Scientists are now further informed regarding pathways which molecules are attributed to with the creation of databases such as *KEGG*, *BioCyc*, *Lipid Maps* and *Reactome* [61], [65]. More recently, fully-automated NMR experimental and spectral profiling has been introduced in order to increase accuracy, precision, efficiency and speed of analysis, for example the use of Bruker's *Biorefecode2* which integrates NMR metabolites into the acquisition software [70]. These processes are all relatively new, and they are continuously developing in this growing

field in order to understand pathogenesis and drug metabolism, and also to harmonise such metabolomics data.

Metabolomics can be described as targeted or non-targeted, dependent on whether specific metabolites are being observed or whether a collection of metabolites are being analysed, respectively. All biological samples have the potential to be used for metabolomics analysis, including blood (plasma, serum, erythrocytes, etc.), CSF, tissue, sweat, muscle and faeces. The focus sample for this study is urine, plasma and CSF. These biofluids are used in a multitude of methodologies in order to understand disease progression and drug pharmacology, as will be demonstrated in later chapters.

Xenobiotics are also a particular focus, since this is an area which genomics and proteomics are unable to explore. Indeed, the biological actions and effects of xenobiotics are able to be 'fingerprinted' via metabolomics investigations, and the metabolism of these can also be explored.

2.1.1 Techniques

As the field of metabolomics has grown, the techniques which are used have advanced tremendously, as has been demonstrated in a review of their growth since the early 1990s [64]; from colour wheels and tastes of biofluids back in the Middle Ages (Figure 2.1) to now utilising techniques such as MS and NMR in order to identify metabolites (both techniques are described in-depth in Sections 2.2 and 2.3) [60]. This combination of complementary techniques is applied in this study in view of their advantages and disadvantages. MS techniques are much more sensitive than NMR ones, i.e. the smallest metabolite and/or drug concentrations that can be quantified with MS are in the pico-molar level range, in comparison to NMR which has a sensitivity at the μM level [59]. In contrast, NMR has the greatest advantage of being non-destructive, and minimal sample preparation time is required [64]. NMR can also simultaneously detect multiple molecules in a sample, without prior separation, which is one of its greatest advantages. However, the cost of the instrument is considerably higher than that of MS, and signal overlap can sometimes be problematic [64], [71]. In addition, MS may not detect ions in view of matrix effects [64]. The exponential growth in the use of these techniques (Figure 2.3 and Figure 2.4) shows the difficulties of choosing between the methodologies and often a combination is preferred.

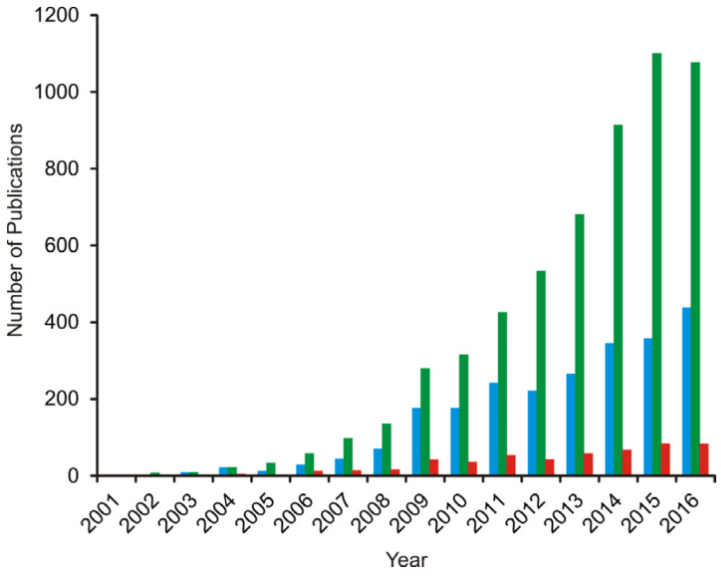


Figure 2.3. Growth in the fields of MS- (green) NMR- (blue) and combined NMR/MS analysis-based (red) metabolomics publications in the research field [64].

There are many variants of MS and other techniques which are widely applied in metabolomics research, and these are commonly described as hyphenated techniques, including LC-MS, Gas Chromatography – Mass Spectrometry (GC-MS), Direct flow Injection Tandem Mass Spectrometry (DFI/LC-MS/MS), Inductively-Coupled Plasma-Mass Spectrometry (ICP-MS), Ultra High Performance Liquid Chromatography-High Resolution Orbitrap Mass Spectrometry (UHPLC-HRMS), Electrospray ionisation mass spectrometry (ESI-MS), or alternatively HPLC with UV or Fluorescence detection (FD) systems [60], [64]. NMR analysis can also be used as a hyphenated technique, namely LC-NMR. A combination of complimentary approaches is likely to increase the overall metabolite coverage, since each technique has some limitations as well as benefits [64]. The wide range of metabolomics instrumentation enables an analyst to acquire different information (dependent on the choice of methodology) in order to determine the nature and concentrations of biomolecules [59]. However, the wide range of techniques does cause difficulty in the harmonisation of data in this field, and the literature expresses a great need to develop all-inclusive reference resources [69] inclusive of research groups such as coordination of standards in metabolomics (COSMOS) specifically developed for this purpose.

Figure 2.4. Challenges of choosing between NMR and LC-MS [72].

Once data has been acquired using one of these 'state-of-the-art' bioanalytical techniques, it is then subjected to multivariate or multidimensional statistical analysis strategies. This methodology can be described as a biological derivative of chemometrics. With many metabolites being identified by a particular technique, such as NMR analysis for example, it is important to know which of these are significant in a particular disease process within the large dataset acquired. Multivariate statistics is employed in order to understand if the datasets between groups of data are significantly different or similar, and if so, which metabolites are contributing to these similarities or differences. The statistics will also provide information on how effective the performance of the tests applied and the significance of the overall result. Molecules or biomolecules with significant differences can be traced back through the cycles which they are derived from in order to further understand the disease or, for that matter, drug mechanism of action or pathway. This chapter aims to provide an in-depth review of the techniques described above, and their uses in metabolomics research approaches are explored throughout.

2.2 Nuclear Magnetic Resonance

NMR is a spectroscopic technique focusing on the observation of the magnetic properties of many different nuclei in different molecules. Unlike Ultraviolet-Visible (UV-Vis) and Infrared (IR) spectroscopies, which focus on the sole principle of Planck's law, NMR relies on unique nuclei which possess magnetic properties in order to be detected. In addition, these nuclei must have an odd number of protons or neutrons, and are therefore NMR-active. Planck's law still applies, but it's significance is secondary, since these molecules are excited by radiofrequencies, but under the influence of a magnetic field. The novelty of NMR

spectroscopy enables the chemist to gain information about molecules at both atomic and molecular levels [73].

The foundations of nuclei possessing a magnetic spin was first recognised by Pauli in 1924, and these ideas were then further developed by Rabi in 1939 who created the first frequencies involving a beam of lithium chloride [74]. Redfield then explored the theory of relaxation processes [75], and NMR was then developed from these principles by Bloch and Purcell from 1945-1970 who won several Nobel Prizes. Continuous wave (CW) methods were used to observe nuclei spin, experiments which use a permanent magnet/electromagnet and a radio frequency oscillator to produce two fields B_0 and B_1 respectively [76]. To produce resonance CW methodologies varied the B_1 field or the B_0 field to achieve resonance [76]. In essence the magnetic field is continuously varied and the peak is recorded on a oscilloscope or an x-y recorder [76]. However, these methodologies advanced, and at present pulse sequences are applied; this methodology is reviewed in more depth below [77].

Characteristic nuclei with magnetic spin obtain a nuclear spin quantum number (*I)* and their own angular momentum (*p*) associated with it, and their relationships are established in equation (1) which is defined for each isotope [78]. This angular momentum (*p*) can be described as that about an axis, and has its own magnetic moment (μ) [78]. The link between these two parameters can be given by the nuclei's gyromagnetic ratio (γ), as in equation (2). These developments are expressed here:

$$(1) \qquad p = I\frac{h}{2\pi}$$

$$(2) \qquad \gamma = \frac{\mu}{p}$$

Every nuclei can be described as having a characteristic magnetic moment (e.g. ^{1}H, ^{13}C, ^{19}F), but in order to be detected by NMR, the correct isotope needs to be observed, i.e. that which has odd number of protons or neutrons in the nucleus [78]. The characteristic nuclei have a spin of $\frac{1}{2}$ and are classified as spin-half isotopes. Within the magnetic field the nuclei have its own magnetic influence too, since it is surrounded by negatively-charged electrons.

Outside a magnetic field, nuclear spins are orientated randomly in space (Figure 2.5, Left). However, once the nuclei are placed in an external magnetic field, the nuclei will orientate with or against that field, once thermal equilibrium is reached [79] (Figure 2.5, Right). The nuclei, with a positive spin, will orientate with (parallel) the field and the nuclei with a negative spin, will orientate against (anti-parallel) that field. This can be described as the *Zeeman Effect*,

which describes this classic orientation of nuclei when exposed to a magnetic field, and this is illustrated in Figure 2.5. The parallel nuclei are at a lower (α) energy state in comparison to those anti-parallel to the field (β). However, almost equal proportions of the nuclei are distributed between both anti-parallel and parallel states, and this is known as the *Boltzmann Distribution* [78]. These ideas are illustrated in Figure 2.5, (3) and (4).

Figure 2.5. The Zeeman Effect: Arrows represent nuclei. Left: Randomly orientated nuclei in space. Right: Magnetised nuclei showing an almost equal proportion of them between the anti-parallel and parallel spin states, if the magnetic field was north, the ascending arrows would be parallel to the field, and the descending arrows would be anti-parallel to the field.

$$(3) \quad \frac{N_\beta}{N_\alpha} = e^{\frac{-\Delta E}{k_B T}}$$

$$(4) \quad E_\alpha = +\frac{1}{2}v_{0,1} \; and \; E_\beta = -\frac{1}{2}v_{0,1}$$

In equation (3), the number of protons in the β and α state are represented by N_β and N_α, respectively. The difference in energy is represented by ΔE, the Boltzmann constant is represented by K_B, and the temperature by T. In equation (4), the energy of β and α are denoted by E_β and E_α respectively and frequency is denoted by ν. When a nucleus is in a homogenous magnetic field, it can be said to align along the z-axis or B_0. In addition, nuclei are said to precess around this axis since complete magnetisation is not possible. This phenomenon is given by the Larmor equation, which defines the angular velocity of the precession movement.

$$(5) \; \omega_0 = \gamma B_0$$

Nuclei in the lower energy state can be transferred to the higher energy state (beta), a phenomenon known as a 'spin-flip' once the nuclei are irradiated with a pulse of radiofrequency. Energy is released when the nucleus 'flips' and can be measured as in normal spectroscopy using equation (5). The higher the radiofrequencies (measured in MHz) are, the more 'spin-flips' occur, which in turn produce a resonance and this is reflected by the

Boltzmann Distribution. The energy, which is absorbed by the nuclei to perform this transition from a low energy state to a high energy state, is expressed by the *Bohr condition*, and this shows the application of Planck's law:

$$(6) \quad \Delta E = h\nu = \frac{\gamma B_0}{2\pi}$$

The relationship between the Bohr condition and the Boltzmann distribution show that a small population of excess nuclei are in the lower energy state [79]. However, this is vital to create an NMR signal, since this small population of nuclei become excited and create the NMR signal. Due to only a small population of nuclei experiencing a change in energy, this renders NMR a relatively insensitive technique relative to MS. This small population difference creates a net magnetisation (M_0) parallel to the magnetic field (B_0).

In order to produce an NMR signal, the magnetisation vector, M_0, must be rotated from the z-axis towards to x or y plane [78]. In order for this to occur, nuclei are irradiated with a radiofrequency pulse orthogonal to B_0, which oscillates at the Larmor frequency achieving the desired resonance [79]. The direction of the nuclei has changed from the z-axis to the x-y-plane. The Larmor frequency can be defined as when the precessing orientations have the same energy as the resonance frequency of the system. This radiofrequency pulse is described as being a 90° pulse, since the magnetisation vector has translated 90°, as shown below in Figure 2.6 by the red arrows. Only when the magnetisation has been transferred to the x,y plane, can an electrical current be generated, which will be detected by the receiver coil, producing a resonance in the x-axis, also shown in Figure 2.6.

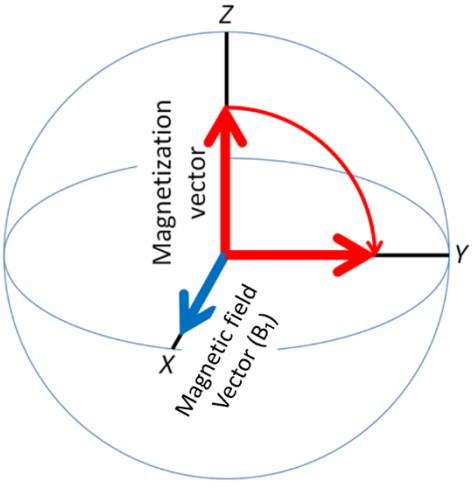


Figure 2.6. Magnetisation transferring from the z-plane (B_0) to the x-y plane enabling an electrical current to be generated [80].

The resonance causes transitions between energy states in a manner which nuclei can be excited to the higher energy state. These are respectively described by equation (4) and Figure

2.5. Only a small quantity of nuclei will move from the lower to the higher energy state, and once in this latter state, this energy, which is absorbed, can induce a voltage, which is then detected by a coil of wire, and which can be amplified. The amplification of the voltage produces a free induction decay (FID), which shows a loss of energy over time as shown in Figure 2.7.

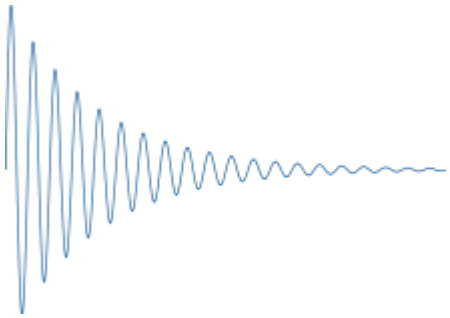


Figure 2.7. Free induction decay (FID), showing an intense magnetisation on the left decaying on the right.

The loss of energy in NMR is ultimately determined by two different parameters known as T1 and T2, which can be described as relaxation. Longitudinal (spin lattice) relaxation (T1) is the re-equilibration of α and β states by transfer of energy to the motion in the sample (lattice). The nuclei flip back to the z-axis, i.e. the spin flips back to non-excited positions, and therefore exponentially decays, similar to a half-life and is a first order process. This is also termed as thermal equilibrium and is important in small molecule NMR spectroscopy. Transverse (spin-spin) relaxation (T2) is the interaction of spins between molecules and unbunching of spins in the x-y plane. Some nuclei precess fast, whilst some nuclei precess slower, dependent on the size of the molecules.

In summary, the novelty of using a magnet ensures that the results are acquired in a different manner to alternative spectroscopic techniques such as UV-Vis and IR as noted above. The energy emitted by the nuclei changing energy states creates energy within the detector coil; this then creates a FID, which arises from a series of cosine waves [78]. These waves are then processed mathematically, using Fourier transform (FT), to create an NMR spectrum [78]. An NMR spectrum is shown as ppm/Hz along the horizontal axis and intensity along the vertical axis. Using ppm is more universal, since it enables the user to compare results arising from all instruments.

In order to analyse NMR spectra, some basic principles need to be outlined in order to understand the information spectra can provide. Ethanol will be used as a typical example of how to analyse a basic NMR spectrum. Ethanol is composed of 2 carbons, 6 hydrogens and

an oxygen which is identical to that of dimethyl ether as shown in Figure 2.8. These principles will outline how the two molecules can be distinguished via NMR spectroscopy.

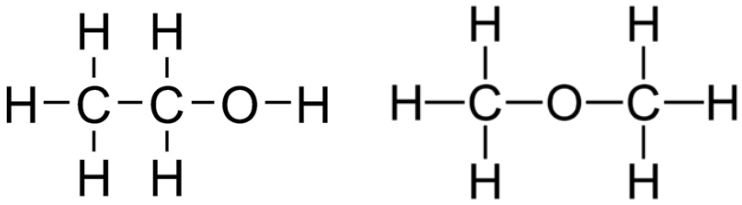


Figure 2.8. Molecular structures of ethanol and dimethyl ether.

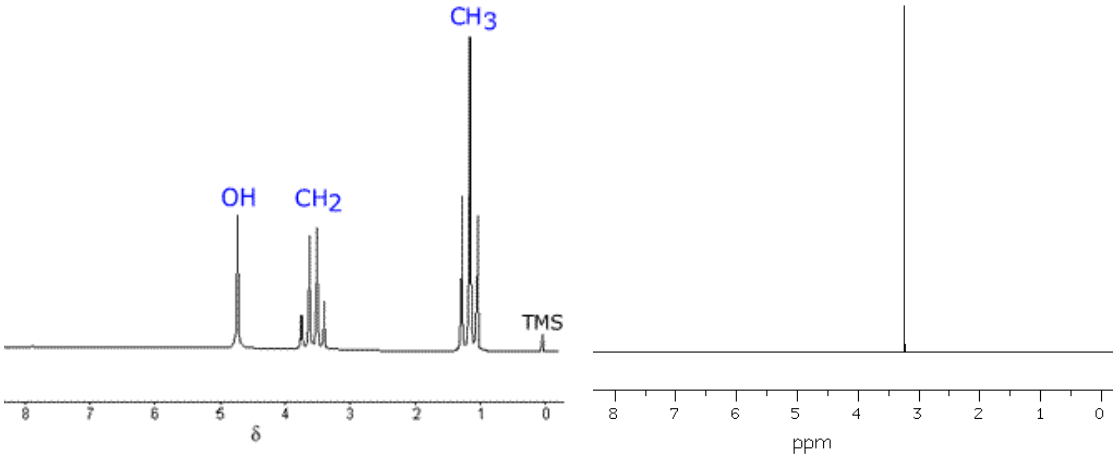


Figure 2.9. ^{1}H NMR spectrum of ethanol (left) and dimethyl ether.

Number of Signals shows the number of hydrogen nuclei chemical environments that are present in the sample - this is dependent on the chemical components of the molecule. *Spin-spin coupling* is also characteristic for high-resolution spectra, and provides information regarding the spin of directly adjacent, surrounding nuclei. The measure of the strength of the interaction between two nuclei is known as the coupling constant. Magnetically-equivalent nuclei do not show any splitting. However, it is important to note that chemically-equivalent nuclei may be magnetically inequivalent. As shown in Figure 2.9, ethanol produce triplet and a quartet signals, which arise from spin-spin coupling. The orientation of the nuclei in their different positions is shown in Figure 2.10. This is also known as *multiplicity*, which follows the (n+1) rule based on Pascal's triangle for simple molecules. This phenomenon occurs due to different energy levels exerting an influence on the signal. The observation of the -CH_3 group producing a triplet is attributable to the neighbouring 2 hydrogens on the adjacent carbon and the different sub-levels within that nucleus. Likewise,

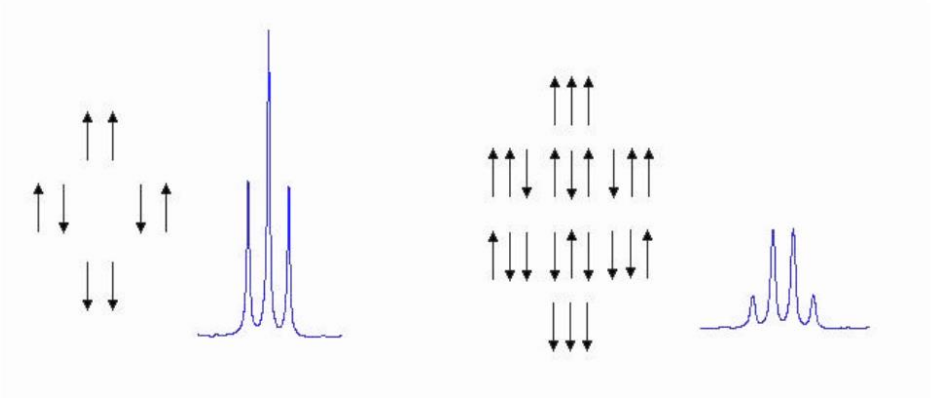


Figure 2.10. Spin-apin coupling of a triplet and quartet showing the different orientations of the nuclei.

the -OH group gives rise to a singlet resonance since the hydrogen will undergo rapid chemical exchange, which removes the potential coupling.

Chemical shift is crucial in NMR; the scale of the NMR is described as being in ppm (or Hz) and shows the location of the signal. The frequency x-axis scale, in ppm, is uniform across all NMR spectrometers, regardless of frequency of the instrument. The precise chemical shift value of a resonance is dependent on what class of nuclei is resonating, and is proportional to the magnetic field that is applied to the nucleus. Below is a guide to ^{1}H NMR chemical shift values (Figure 2.11), which illustrate basic ideas regarding expected shifts. Hydrogens bonded to more electronegative nuclei (e.g. those within carboxylic acids, aromatic molecules, etc.) tend to resonate downfield, whereas less electronegative nuclei tend to resonate upfield (e.g., alkanes). Such chemical shift phenomena are present because the electrons around the nuclei have a shielding effect against the applied magnetic field, and the nucleus effectively becomes magnetically shielded or deshielded. This can be represented by the *Larmor* Frequency.

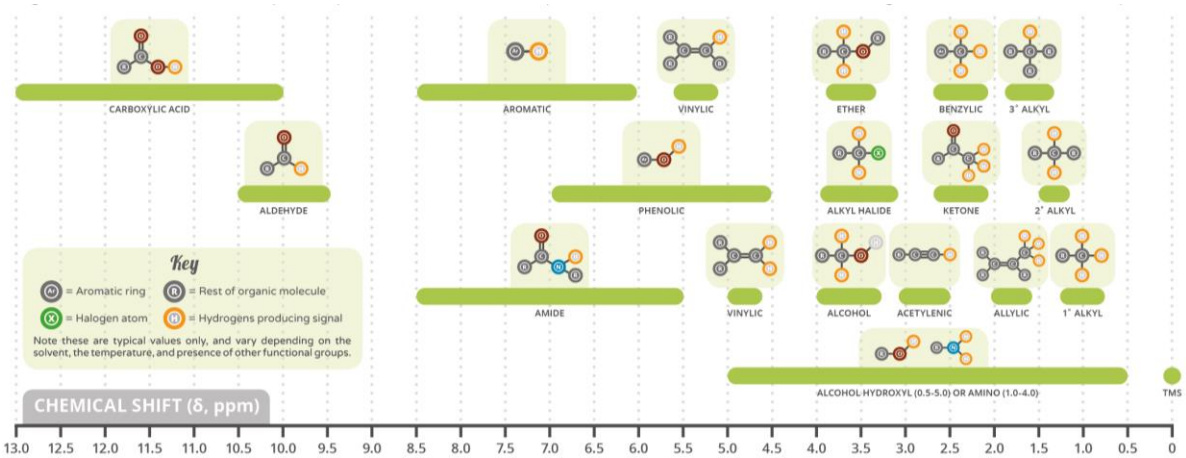


Figure 2.11. Simple guide to NMR chemical shift values [81].

Integration of NMR spectra enables the user to quantify the concentration of molecules from spectra. This information can be extracted from the intensity of the signal, but it is also vital to account for the number of hydrogens attached to them producing this signal. Integration involves analysis of the area underneath resonances. The NMR spectra therefore provides quantitative information regarding hundreds of metabolites in the case of biofluids.

This is an outline of the very basic rules of NMR spectroscopy and takes into account the principles for most first-order systems, also known as AX, AMX and A_2MX systems in which hydrogen nuclei are chemically- and therefore magnetically- equivalent. However, the coupling of nuclei and how they interact with one another becomes more complex with larger molecules with different types of nuclei, which are not chemically or magnetically equivalent, but share the same three-dimensional 'space'. Nuclei themselves are small magnets, and once placed in a magnetic field, their own magnetic field affects surrounding nuclei which 'disobey' these basic rules and create different splitting patterns outside of the n+1 rule. More complex formations of resonances can occur, and these are created from non-first order systems, such as AA'XX' AB and ABX ones.

Nuclei which are the focus for this thesis are primarily ^{13}C and ^{1}H, since biofluids are predominately composed of organic molecules, as indeed are most drugs. However, NMR is not limited to observing ^{13}C and ^{1}H; metal ions can also be analysed as part of metabolomic analysis, using chelating agents such as Ethylenediaminetetraacetic acid (EDTA). Furthermore, phosphorus (^{31}P) and fluorine (^{19}F) NMR probes can also be utilised. All these nuclei possess a nuclear spin quantum number (I) of $\frac{1}{2}$ which can both be analysed by NMR spectroscopy. It is easy to elucidate metabolites with the use of NMR, however; more specialised techniques such as LC-MS or LC-MS/MS are required when observing some molecules, if they are present at only limited quantities in biological fluids and this will be explored in the following chapters.

The main advantages of NMR for metabolomic analysis is that it provides a non-targeted metabolomic, near global approach, which is able to detect a combination of hydrophilic and lipophilic metabolites, dependent on pulse sequence and extraction methodologies applied [68]. Furthermore, samples do not require pre-preparation in most cases, and the quantitative analysis is fast, non-invasive, non-destructive and reproducible [68]. In addition, only small sample quantities are required, which overall renders NMR as the methodology of choice.

The following sections will review specific NMR pulse sequences which are applied to biological fluids throughout the practical chapters. Biological fluids require different experimental and acquisitional conditions in order to produce desired results, since the

biomolecules present are generally of limited quantities. Larger molecules could 'hide' the smaller ones which would impact analysis, so overcoming this obstacle is further explored. Furthermore, this section aims to provide an insight into techniques which enable the identification of such molecules, which is of much importance when distinguishing biomarkers.

2.2.1 Water Suppression

Saturation of the water signal is particularly important for biomedical NMR analysis, especially when considering biofluids. Some biological samples, such as urine, contain up to 95% (w/w) water, in addition to many thousands of biomolecules [60]. However, the presence of such a high quantity of water is troublesome to spectroscopists, since it hides valuable information from surrounding metabolites which resonate at a similar frequency. This is problematic, as small concentrations of metabolites could be missed entirely, and so techniques have been developed in order to overcome this issue. Without resolution of this problem, such metabolites would be masked by the overlapping solvent signal.

A standard pulse sequence consists of using a 90° pulse and data acquisition, the FID is recorded and then FT is applied to produce the spectrum. However, for water suppression overcoming this issue requires a change in pulse sequence to minimise the signal. The most employed NMR techniques applied for water suppression in metabolomics are NOEPR (Nuclear Overhauser Effect pulse with Presaturation during relaxation and mixing time) and PRESAT (Presaturation). These techniques are usually used succinctly as a NOESY-PRESAT and provides an untargeted metabolomics approach and in combination aim to enhance the signal of the small metabolites whilst eliminating the solvent signal.

The PRESAT experiment is composed of two pulses with a low radiofrequency which selectively saturate the water signal by irradiating during the relaxation delay. These long pulses targeted at the order of solvent T1 ensure that the nuclei are not allowed to relax and therefore a signal, in turn, is suppressed. If the nuclei do not relax, they cannot be detected via NMR. Any group which chemically interchanges with water such as molecules with -OH, -NH groups will be suppressed too. NOEPR focuses on reducing the solvent signal during the recycle delay and mixing time as the pulse sequence is similar to the NOESY experimental. The sequences are structured below:

$$(7)\ PRESAT\ [Presat(RD) - 90°]$$

PRESAT is where the duration of the pulse is inversely proportional to the width of the area excited which in turn means the longer the pulse signal, the narrower the spectral bandwidth.

$$(8)\ NOEPR[-90° - T1 - 90°tm - 90]$$

NOEPR is where T1 is a short delay and tm is the mixing time in which the water is specifically irradiated which reduces the signal by enhancing the population difference in nearby nuclei [82].

These techniques are utilised together throughout this thesis in order to gain the results acquired in a manner which is simple and robust [78]. Although other water suppression techniques can be utilised such as Water suppression enhanced through T1 effect. WATERGATE and Excitation pulse sequences are other methodologies which can be utilised for water suppression.

2.2.2 Carr-Purcell-Meiboom-Gill (CPMG) Pulse Sequence

CPMG was first developed by Carr and Purcell and later refined by Meiboom and Gill. CPMG pulse sequence is a ^{1}H NMR which is specifically used for pre-extracted plasma in this study, in order to reduce the signals occurring from proteins and lipids [82]. Proteins and lipids produce a large baseline and the CPMG pulse sequences aims to reduce this. The pulse sequence focuses on the small molecules, which are of most interest in untargeted metabolomics studies [83]. The larger molecules are allowed to relax first, as they have small T2 values, prior to detecting the small molecule resonances [83].

For the CPMG pulse sequence, a standard 90° pulse is applied and the maximum magnetisation is reached and over time the magnetisation decays. The loss of total magnetisation is a result of spin-spin relaxation and each nuclei obtain their own magnetic field. In turn, nuclei precess at different Larmor frequencies. Another 180° pulse is applied after a period of time (τ), which enables individual nuclei to be refocused and attain homogenous magnetisation.

(9) CPMG $[90^{o}_{x}$ - τ - 180^{o}_{y} - τ - 1st echo - τ - 180^{o}_{y} - τ - 2nd echo - … $(\tau - 180^{o}_{y} - \tau)_{n}$

2.2.3 Total Correlation Spectroscopy

1D TOCSY can be performed as a targeted metabolomic approach in order to gain more information regarding the molecules structure and bonds. This technique shows relayed

correlations between nuclei which are up to 2-3 hydrogen bonds apart, sometimes 4 bonds apart [78]. If a nucleus is excited, other nuclei are excited through spin-spin coupling which produces resonance. For more extensive information regarding molecular structure 2D NMR techniques are usually employed as these provide a larger picture of the entire sample.

2.2.4 2D NMR

The use of 1D NMR alone sometimes cannot provide enough information regarding small molecules and their bonds. 2D NMR is often used in conjunction in order to investigate and confirm molecular structures which reinforce the specificity of the technique[84]. 2D NMR Spectroscopy operates by excitation of nuclei at particular frequencies in order to correlate those with alternate nuclei more than three bonds away. This gives a bigger picture of how a molecular structure is composed. 2D NMR plots across horizontal and vertical axes, the two different axes can be two different nuclei (heteronuclear) or two of the same nuclei (homonuclear), dependent on the experiment chosen. The key difference between 1D NMR and 2D NMR is that they are plotted by two frequencies instead of one frequency defined as F1 and F2. A cross-peak is created by a co-ordinate style plot. The coordinates which are set along the diagonal are a connection of the same molecule. The NMR needs to record a 1D NMR spectrum prior to the 2D NMR spectrum in order for this to be loaded into the frequency domain. The 2D experiments follow the general scheme of 4 stages (Figure 2.12).

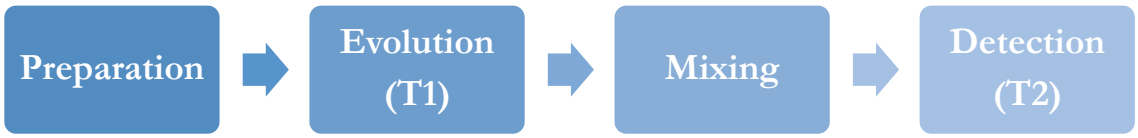


Figure 2.12. 2D Experiments Scheme.

The main goal of 2D NMR techniques is to overcome issues regarding overlapping resonances to confirm molecular identity [68]. The main techniques which are going to be utilised include 2D COSY, which observes nuclei through H-H correlations (coupled protons) and is able to monitor relationships between hydrogens 2-3 bonds apart and 2D TOCSY to observe all correlations.

2.3 Liquid Chromatography-Mass Spectrometry

Some molecules cannot be detected by NMR and more sensitivity is required in order to view these molecules in limited concentrations. When this is the case, LC-MS is a hybrid technique which is aimed at the polarity and molecular weight of molecules. The theory of HPLC had been initially developed in 1941 by A. J. P. Martin and R. L. M. Synge [85] and it was not until over 25 years later in 1966 C. Horvath and S. Lipsky utilised HPLC as a technique [86]. In contrast, the roots of MS were established much earlier, Prouts hypothesis recognised the significance of molecular weights in the early 19th Century [87]. However, it was not until 1919 the first mass spectrometer was built by F. W. Aston [88]. It would also be important to note in 1918 prior to F.W. Aston, A.J. Dempster reported on his mass spectrometer and established the theory and design, but his discovery was used for applications to allow the development of the atomic bomb and nuclear power [89]. This contrasts hugely with the work of F.W. Aston who primarily focused on different isotopes and developed the whole number rule, he ended up becoming a fellow of the Royal Society and received a Nobel Prize in Chemistry.

The combination of using LC and MS together as a technique was first developed in 1968 by Victor Tal'Rose and was further developed by Hewlett Packard in 1979 [90]. This had come a long way from initial theories in the 19th Century and the progress over a few hundred years was a rather large leap.

The LC component is a separation technique, dependent on the column used, LC will separate molecules by the use of a mobile phase and a stationary phase. If a polar column is used, molecules which are polar in the sample will 'stick' to the column. In contrast if a non-polar column is used, molecular which are non-polar will 'stick' to the column. The principle of the technique is based on *Van De Waals* forces of interaction between molecules. HPLC contrasts with LC due to the high pressure being exerted onto the sample which ranges from 50-350 bar whereas, LC alone relies on the force of gravity. The length of the column effects the resolution of the results, the longer the column the better the resolution, but other parameters such as flow rate, temperature are also important.

The use of different phases in HPLC are categorised into two types normal-phase and reversed phase [91]. Normal phase HPLC uses the polar column and the column is typically packed with silica particles and a non-polar solvent is utilised [91]. Reversed-phase HPLC uses the non-polar column and the column is packed with silica particles too however, they typically have long hydrocarbon chains attached to make the column non-polar [91]. For reversed phase HPLC a polar solvent is usually utilised [91]. Reversed-phase is typically

utilised in these studies as the molecules found in urine are typically polar and therefore will pass through the column more quickly.

After separation, the sample can be passed through a mass spectrometer which will bombards the molecules with electrons in order to create ions. The liquid is transformed into an aerosol in order to be suitable for mass spectroscopy analysis. The ions are then detected and various detectors are used. In the case of quadrupole, the ions travel through 4 quadrupoles and dependent on the ionic strength of the ion, different ions will be detected at different stages. There are positive and negative ions produced as part of this process and the mass spectrometer can be tuned to observe negative ions or positive ions. The mass to charge ratio of the ions can be associated with different molecules as these are characteristic to a particular structure which are useful for building up a compound. Mass spectrometry also provides information regarding molecular weights too and a trace of the overall molecular weight of a target compound can typically be seen.

The mass spectrometer operates in two different modes SCAN mode and SIM mode. The SCAN mode observes the entire array of molecules which are in the solution whereas the SIM mode will look for a target molecule. There are advantages and disadvantages of the two different modes, SIM mode is more sensitive and useful for quantification purposes, whereas SCAN mode takes into account the entire sample composition.

2.3.1 Preparation Techniques

Although NMR and MS are highly multicomponent techniques and some samples need little or no preparation prior to analysis, on occasion, some samples do require preparation. The preparation of a sample may be necessary in order to concentrate the sample, in order to get more information regarding the content of the sample or it may involve splitting the contents of the samples into lipophilic and hydrophilic species. Different preparation techniques have advantages and disadvantages and some used in this thesis are outlined below.

Solid Phase Extraction

Solid Phase Extraction (SPE) is a technique which is often used in conjunction with both NMR and LC-MS and is an excellent preparation technique which removes compounds in liquids dependent on the chemical properties. It functions similarly to HPLC by pressing the liquid (mobile phase) through a syringe within a solid (stationary phase) which either has a hydrophilic or hydrophobic surface, enabling certain molecules to pass through.

Syringe Filtering
Syringe filtering is useful for removal of matter prior to LC based methods and is composed of a circular filter and a syringe which combined will let a liquid pass through however, particles will not. The size of the membrane within the circular filter determines what size molecules will pass through.

Freeze Drying
Samples can be freeze dried in order to remove water from the sample so the sample can be reconstituted into a different solvent, or concentrated. This process is completed by rapid freezing under a vacuum which removes ice *via* sublimation. The molecules that remain can be reconstituted into a solvent which is more lipophilic such as chloroform or methanol which can provide further information.

Centrifuge
Centrifuging is another methodology to remove solid matter from the sample. This works by rotating the samples at a high speed which causes the liquid matter to suspend and the solid matter to move down to the centre. The liquid supernatant can then be extracted and used for analysis. In biofluid NMR analysis, removal of solids from cellular matter ensures shimming is optimal.

2.4 Multivariate Statistics

When handling colossal and complex datasets from untargeted metabolomic approaches, which include a plethora of metabolites at a range of concentrations, which are generated from NMR, it can be difficult to interpret and understand the significance of the data acquired. Multivariate statistics have integrated into analysis of NMR, due to the number of data points produced from the output of analysing complex biofluids. Multivariate (MV) statistics aids in dealing with large datasets into visual formats which are more comprehensive for the analyst by recognition of patterns in the data. The combination of using a scientific technique alongside statistics in this manner is termed chemometrics. The following section explores some of these visual formats and their uses in aiding interpret the complex datasets. Prior to use of these techniques it is important the data is prepared using normalisation, referencing and scaling these will be described in more depth in later chapters. All univariate and MV forms of statistical analysis will be performed using *XL STAT 2016*, *MetaboAnalyst 4.0*, *ROCCET* and *MetATT* software.

2.4.1 Principle Component Analysis

Principle Component Analysis (PCA) is arguably as the most widely applied and useful multivariate statistic for metabolomics [62]. The main aims of PCA are to establish patterns in large datasets concentrating on similarities and differences in the data via PC 'clusterings' and enables to the user to detect outliers which are statistically significant [84]. The data will cluster and the classification of each of the groups can be evaluated [84]. It is a unsupervised tool and therefore unbiased which is advantageous in non-targeted metabolomic studies [92]. Thus, the technique aims to find the maximum variance in a data set (X) without considering the class (Y) [92]. PCA operates by examination of the interrelations among a set of variables in order to identify a new structure of those variables namely principle components (PCs) or scores which have a weight or loading [84], [92]. PCA performs a linear transformation of the starting variables with appropriate weighting coefficients and the tool attempts to retain as much variance as possible with each PC remaining uncorrelated with each other [62], [84]. The primary component in the PCA will retain the greatest variance and subsequent components will contain less and less until there is no variance [84].

From the data matrix given, usually an NMR dataset matrix is created in the case of metabolomics, eigenvectors and variance are established and utilised to produce the PCA plot. The eigenvectors are formed of the non-singular portion of the sample covariance matrix (S) for the column. The eigenvalue will reflect the number of variables associated with the principle component. For metabolomics, a set of metabolites which are negatively or positively correlated will form a PC. A second set of metabolites which are correlated will form a second PC and PCA will compare these two sets of loadings.

2.4.2 Partial Least Squares-Discriminant Analysis and Orthogonal Partial Least Squares Discriminant Analysis

Partial Least Squares Discriminant Analysis (PLS-DA) previously nominated as Projection to Latent Structures is another statistical tool commonly used in metabolomic analysis however, this tool is supervised unlike PCA and considers the class memberships [92]. PLS-DA optimises the separation between different groups of samples, by maximising the covariance between independent variables and discovering a new linear subspace between explanatory variables [59]. The advantages of this approach is that the technique is able to handle highly collinear and noisy data which is attributed to NMR and LC-MS outputs [59].

PLS-DA enables complex datasets to be visualised through easily interpretable scores plot that can show separation between groups [59].

This technique does not impose orthogonality and requires X scores to be uncorrelated, closer to the PLS regression. PLS creates scores and loadings described as latent structures/variables which summarise the set of X and Y values. The production of latent variables by this method consist of linear combinations of the original variance that preserve as much between X and Y as possible.

The problem with PLS-DA is that variation which is not directly correlated with the variables is still present in the scores which is problematic when the number of classes increases. In contrast, Orthogonal Partial Least Squares Discriminant Analysis (O-PLS-DA) addresses these issues by incorporation of an Orthogonal Signal Correction filter into the original model which separates the predictive variation from the uncorrelated variation.

PLS is a supervised technique which takes into consideration the groups and can potentially overfit if the dataset is small and cross validation is required in order to confirm the strength of the test. A permutation test is usually performed to ensure the test is employed correctly testing the multivariate independence of disease classification and prognostic stratification groups [84]. A Q^2 value gives the accuracy of the test.

Variable importance in projection (VIP) score is a weighted sum of squares of the PLS loadings and will show the regions causing the most significance in the correlations or lack of correlation between the PCA score [92].

2.4.3 Computational Intelligence (neural networks)

Computational Intelligence techniques, including Optimisation Algorithms and Machine Learning Algorithms (supervised and unsupervised neural networks) can be employed in order to recognise metabolic classification patterns for each of the disease and disease severity classifications [93]. In addition, reinforcement learning algorithms can be used to identify classification and have been previously applied to the medical research domain [93].

Optimised algorithms will be used to select the most discriminatory biomarker variables from the datasets for use in constructing machine learning classification models. Supervised learning techniques will then classify the diseases. Feed forward and cascade forward networks can be formed trained with backpropagation in order to classify disease groups [94]. Different algorithms have been previously applied in metabolomics including, but not

limited to Levenberg-Marquardt, Bayesian regularization, Broyden–Fletcher–Goldfarb–Shanno Quasi-Newton, Resilient, scaled conjugate gradient, conjugate gradient with Powell/Beale restarts, conjugate gradient with Fletcher-Powell, conjugate gradient with Polak-Ribiére, one step secant, variable learning rate gradient descent, gradient descent with momentum and gradient descent [95]. These algorithms can be multilayer or single layer networks [95]. Deep learning is being used more frequently to observe metabolomics data; a feed forward network has shown promise in distinguishing estrogen receptor positive and negative breast cancer tissue samples for example.

2.4.4 Support Vector Machines

Support vector machines (SVM) is also used for classification purposes and indeed is a supervised technique. The SVM classification uses non-linear decision functions in the input space in which both a higher dimensional feature space is mapped [92]. Features are separated by a maximum margin hyper plane [92]. Cross validation error rates are applied to features which are based on their relative contribution in the classification [92]. Non-important features are removed and thus a series of SVM models are created [92]. The best ranking model can be used and is ranked by its frequency [92].

2.4.5 Self-Organising Maps

Self-Organising Maps (SOMs) are utilised in order to explore self-similarities between the spectra acquired and hence clustering's - arising from each source of variation involved [84]. SOM's are unsupervised neural networks using a grid method which consists of nodes [92]. The model is trained to represent subsets of data, as the model clusters random values and the user specifies the expected dimension of the grid [92]. Aggravated expression profiles are then formed from the k-means and SOM which are presented in the form of a line graph on top of each other using feature values [92].

2.4.6 Heatmaps

The concept of heatmaps is dated back to Loua in 1873 who shaded in a matrix which was drawn and coloured by hand [96]. Cluster matrices were then developed in 1914 by Brinton which reveal more structure [96]. The advances since then are vast, many computer software

include this statistical tool and its applications are recognised in various fields, from accounting to science.

Heatmaps are a visual tool which are applied to represent data, which will be put into a matrix and colours will reflect the values in a hierarchical cluster structure. Heatmaps appear as colourful tiles on a rectangle matrix and hierarchical cluster trees are marked on the margins. The creative history of the heatmap is reflected in its presentation and is a crucial tool for MV statistics. Colours usually reflect patterns of increases and decreases in metabolites across the matrix whereby red represents increase and blue represents decreases. The intensity of the colour will also depict the quantity of increases or decreases whereby, a more neutral red would indicate a slight increase, whereas a strong red would indicate a strong increase.

Heatmaps are used in metabolomics in order to distinguish and increase or decrease in metabolites across the samples relative to each other and are an extremely useful visual tool.

2.4.7 Classification Trees

Classification trees are able to construct non-linear relationships and are subcategorised into adaboost, bagging or random Forest (RF). RF's generate many decision trees which are constructed by randomly selected variables from a bootstrap sample at each branch and were first introduced by Brieman [59], [92]. One third of the bootstrap samples are left out during the construction of the trees for the purpose of being used as a test sample, in order to obtain an unbiased estimate of the classification (OOB) out-of-bag error [92]. Variable importance can be measured by increasing the OOB error [92].

This technique can also handle large datasets, measure predictor variables, a measure of internal structure of the data and also able to handle missing values [59]. RF produces an internal unbiased estimation of the generalisation error and is robust to outliers as described and is also supervised like PLS-DA [59], [92].

2.5 Summary

Overall, metabolomics is a growing field and gives an insight into the pathology of different biological systems which can aid in personalised patient care, diagnosis, therapy and overall outlook of patients. The versatility of the subject enables research to also delve into discovery

of potential biomarkers or characteristic features in diseased patients. Therefore, it is crucial to expand the research in this particular field as the information it may provide is profound. MVA ensures that the large data quantities are manageable and understandable. This combination of techniques and technology are at the forefront of chemometrics and metabolomics. A final figure, (Figure 2.13), shows the work flow of metabolomics data from acquisition to results and encapsulates the experimental design.

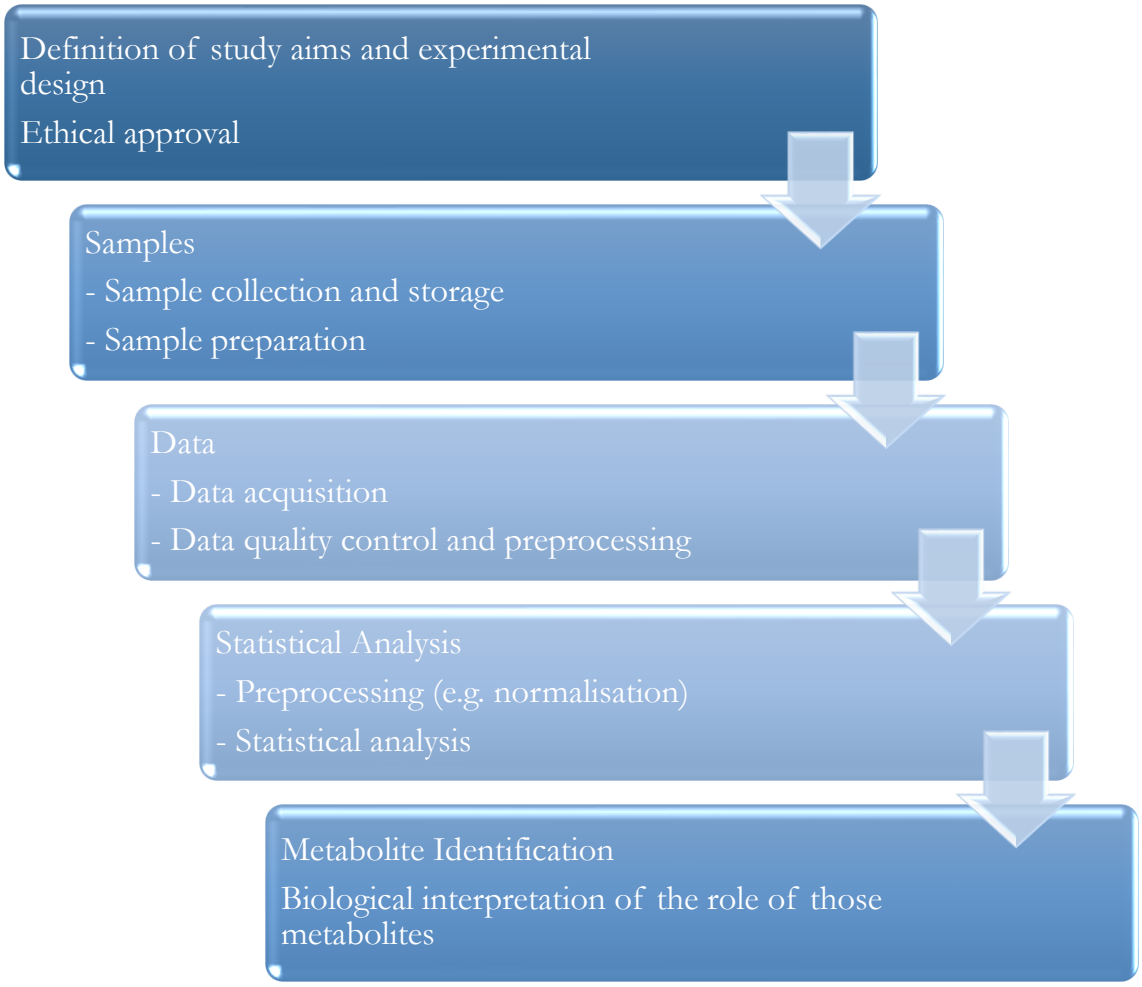


Figure 2.13. Metabolomics Experimental Design ideas adapted by [1].

Chapter 3: ^{1}H NMR-Linked Metabolomics Analysis of Urine and Blood Plasma Collected from Gaucher's Disease Type 1 Patients: Challenges Associated with Xenobiotics and Experimental Design

3.0 Abstract

Background: ^{1}H NMR metabolomics provides potential scope as a non-invasive approach to discover candidate biomarkers for, and further understanding of metabolic pathway perturbations in Gaucher's Disease Type 1 (GD1). GD1 is a complex multisymptomatic disorder affecting multiple organs, including the liver, spleen, skeleton and brain. In order to further understand GD1 and assess the effectiveness of specific treatments for GD1 (imiglucerase, velaglucerase alfa and eliglustat) a metabolomics approach was utilised.

Methodology: Age- and sex-matched plasma and urine samples were provided from fasted and unfasted controls, and unfasted GD1 participants. ^{1}H NMR datasets were acquired using a 700 MHz cryoprobe facility. ^{1}H NMR spectral data was processed, integrated and then prepared for multivariate metabolomics analysis using normalisation and scaling techniques. Since the fasting status unfortunately differed between the two cohorts, so all metabolites known to be affected by fasting status, such as glucose etc., were removed from the dataset prior to statistical analysis. Drugs such as acetaminophen, ibuprofen and acetylsalicylic acid were taken by the GD1 cohort which also resulted in removal of many ^{1}H NMR bucket regions from the urinary ^{1}H NMR profiles. However, no xenobiotic signals in plasma spectra acquired were detectable.

Results and Discussion: No significant differences between the urinary profiles of GD1 and control participants were found. However, some significant metabolic differences in their plasma samples, including downregulated high-density- and low-density-lipoprotein triacylglercol, and upregulated lactate, 3-D-hydroxybutyrate and acetone levels were observed. The influence of the differences in fasting status and potential biochemistry changes due to the diseases status of the participants is discussed.

Conclusions: ^{1}H NMR-linked metabolomics analysis can successfully distinguish between GD1 and control plasma profiles, and these results provide insight regarding new metabolomic markers for GD1. Further work could aim to validate the findings by the synchronous observation of both untreated and treated GD1 participants in order to confirm findings. Alternatively, an animal model system could also be employed in order to retain

consistency in the experimental design and reduce variability introduced from xenobiotics such as drugs or dietary supplements.

Keywords: Nuclear Magnetic Resonance; Metabolomics; Lysosomal Storage Diseases; Gaucher's Disease; Blood Plasma; Biomarkers

3.1 Introduction

In this pilot investigation, conventional ^{1}H NMR metabolomics approaches were employed in order to identify new, and potentially valuable, biomarkers for Gaucher's Disease Type 1 (GD1) in human blood plasma and urine, particularly those which correlate with disease severity and progression. Such an investigation may provide further understanding of GD1 pathogenesis through the identification of defects in key metabolic pathways involved. This ^{1}H NMR-based strategy has been supplemented by the application of newly-developed metabolic pathway and quantitative metabolite set enrichment analyses. The ^{1}H NMR dataset matrix was also interpreted using both multivariate and univariate statistical approaches in order to further understand dataset variation and its contribution towards validation of potential biomarkers. These software-based approaches served to provide much valuable information regarding biomolecular connectivities, and hence the biological/metabolic pathways involved in the pathogenesis of GD1. This methodology also permitted the recognition of metabolic 'signatures' that may enhance our understanding of the pathogenic mechanisms of GD1 at the molecular, biochemical and cellular levels. Furthermore, such metabolomic profiling has provided valuable insights into the effectiveness of enzyme therapy, as a current treatment.

3.2 Specific Aims

- To observe the metabolic profiles and patterns of blood plasma and urine collected from GD1 patients, and compare these with those of healthy controls with the outlook of improving the diagnosis of these patients; these studies will involve the seeking and identification of candidate biomarkers which may be useful for the future monitoring of clinical progression.

- To further understand perturbations in and disturbed metabolic pathways in GD1, with a particular focus on identifying how such modifications may be involved in its chemopathological mechanisms.
- To further understand the impact of treatment outcomes of enzyme replacement and substrate reduction therapies on GD1 metabolome.

3.3 Materials and Methods

3.3.1 Approval for Research

All control samples were collected with informed consent and approved by the appropriate Research Ethics Committee at De Montfort University (DMU) in accordance with the declaration of Helsinki 1975 (revised 1983). Samples from GD patients were provided by the Department of Pharmacology, University of Oxford.

3.3.2 Human Plasma Collection

Twenty-two control participants were recruited directly from DMU who were fasted for a 12-hour period, whilst 28 participants with GD1 (4 further control participants were recruited by Prof. Platt's team at the University of Oxford). Blood samples were collected by a fully trained phlebotomist *via* venepuncture on-site at DMU, and all blood samples were collected in lithium heparin collection tubes. Whole blood samples were centrifuged for 15 min. at 4,300 rpm and 4°C on-site and the plasma was frozen at -80°C and ready for ^{1}H NMR analysis.

Control participants were provided with a participant information sheet and signed a consent form. Control participants also provided information such as diet in the form of a food diary, demographic information such as BMI, along with their use of supplements and medications. All samples were anonymised on collection and numerically coded. Blood plasma was collected into heparinised tubes in order to avoid NMR analytical complications arising from the use of those containing EDTA or citrate anticoagulants. The complications of adding an EDTA anticoagulant include its complexation of Ca^{2+} and Mg^{2+} ions, and this has been demonstrated in several works elsewhere [26], [97]. Samples from the University of Oxford were provided by participants who were not fasted but were collected, centrifuged and stored in the same manner.

Class	Gaucher Disease	Control	Total
Mean Age	54.17	39.88	47.29
Range (Years)	33-77	21-64	21-77
Males	12	16	28
Females	16	10	26
Total	28	26	54

Table 3.0. Breakdown of age and genders of healthy control, and GD1 (*OMIM #230800*) plasma and urine samples.

Samples collected represent those from a near aged-matched cohort with a similar ratio of females to males. In the GD group, all except one sample were GD1 with the remaining sample being Type 3 GD (GD3), which is excluded from this table (*OMIM #231000*).

3.3.3 Human Urine Collection

Nineteen fasted control participants were recruited directly from DMU. An additional 29 unfasted ones with GD1, along with 4 unfasted control participants were recruited by the University of Oxford. Urine samples from DMU were collected by participants after ~12 hour fasting period, was the first pass of the day and midstream urine collection. They were collected onsite at DMU in sterile universal containers and stored at -80°C until ready for analysis. Control participants provided the same information as described above in the blood plasma section 3.3.2. Samples were collected in the same manner at the University of Oxford, although samples were collected from unfasted participants. The urine provided was from samples time-matched to the plasma ones. However, some participants did not provide a urine sample.

Class	Gaucher Disease	Control	Total
Mean Age	55.27	44.17	50.36
Range (Years)	33-77	31-64	31-77
Males	12	8	20
Females	17	15	32
Total	29	23	52

Table 3.1. Breakdown of age and genders of healthy control and GD1 urine samples.

3.4 ^{1}H NMR Analysis

3.4.1 ^{1}H NMR Sample Preparation of Human Plasma

Samples were prepared using 450 µL of plasma, 50 µL of sodium azide, 50 µL of phosphate buffer (pH 7.0) and 50 µL of D_2O containing no added 3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid, sodium salt (TSP) internal standard. Samples were analysed at the University of Oxford using the Bruker AVII 700 MHz NMR equipped with a ^{1}H TCI cryoprobe (Chemistry Department, University of Oxford) at an operating frequency of 699.989 MHz and a probe temperature of 298 K. Spectral acquisitions involved the collection of 65536 data points using 32 scans across a spectral width of 15.99 ppm. For the plasma, the CPMG sequences was used which is also previously described in Chapter 2, section 2.2.2) in order to suppress the broad protein signal envelope which obscures the visibility of many low-molecular-mass metabolite resonances (Appendix 1). Samples were loaded onto an automated sample belt using a sample changer.

3.4.2 Plasma NMR Data Preprocessing

All samples were referenced and aligned to alanine -CH_3 at δ = 1.47 ppm, baseline corrected and phased. Once this was completed, samples were then bucketed using *ACD Labs 12.0*. Manual bucketing requires the analyst to observe NMR signals visually and bucket accordingly, with prior knowledge of signal multiplicity. This is an important step since ^{1}H NMR chemical shifts can vary, and depend on temperature, pH and ionic strength. It was noted that histidine signals shifted slightly in the spectra, probably a consequence of small variations in pH, and therefore manual bucketing was considered to be the most appropriate method. Uniform bucketing was not performed, since this can lead to further problems such as signal splitting, which would have been the case for histidine. Intelligent bucketing can also be performed; however, this can additionally give rise to split signals. The raw integral buckets were then imported into Microsoft™ Excel to create ^{1}H NMR data matrices. Any buckets attributable to baseline noise or water were removed prior to multivariate and univariate statistical analyses.

3.4.3 ^{1}H NMR Sample Preparation of Human Urine

Samples were analysed using the same facility described above with the same operating frequency and probe temperature. Spectra acquired collected 65,536 data points using 32 scans across a spectral width of 15.99 ppm. For the urine samples, NOSEY Presaturation (*noesygppr1d)* was used as previously described (Chapter 2, section 2.2.1) targeting the intense water signal located at δ = 4.70 ppm. Samples were loaded onto an automated sample belt using a sample changer.

3.5.4 Urine NMR Data Preprocessing

All samples were baseline-corrected, phased and aligned to TSP. Once this was complete, spectra acquired were then bucketed using *ACD Labs 12.0* using two different methods, manual bucketing and intelligent bucketing. For intelligent bucketing, buckets had a width of 0.02 ppm and 50% width looseness. The buckets were then imported into Microsoft Excel and the data was normalised to TSP. Any buckets attributed to baseline noise, drugs and their metabolites, supplements, urea (δ = 5.80 ppm) or water (δ = 4.70 ppm) were removed.

3.5 Results

3.5.1 ^{1}H NMR Plasma Profile Results

^{1}H NMR profiles were assigned using the *Bruker Topspin 4.0.2* software module; spectra acquired were referenced to alanine-CH_3 (δ = 1.47 ppm), overlaid and labelled (Figures 3.0-3.2). The water signal located at δ = 4.70 ppm was eliminated from the figures. This process used articles [61], [98]–[100] and HMDB in order to confirm the assignments of signals from their chemical shift values, coupling patterns and coupling constants. No assignments were attributed to drugs ingested in the plasma profiles, despite GD1 participants being on various medications, because the drugs are in concentrations too low for detection by ^{1}H NMR.

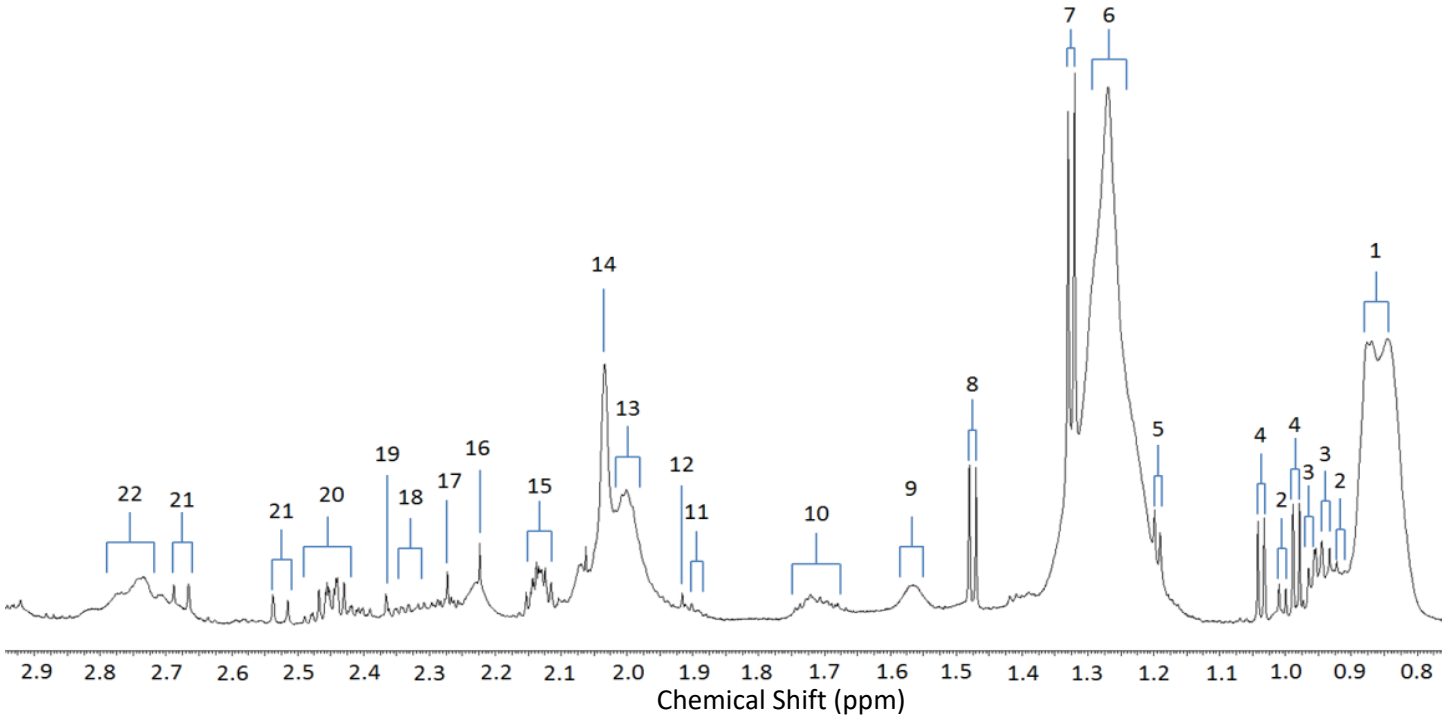


Figure 3.0. Labelled 700 MHz CPMG ^{1}H NMR spectra of GD plasma sample from 0.75 ppm-2.95 ppm; [1] very low density lipoproteins (vLDL)/ low density lipoproteins (LDL) / Triacylglycerol (TAG)-C$\underline{H}_3$; [2] Iso-Leucine-C$\underline{H}_3$; [3] Leucine-C$\underline{H}_3$ [4] Valine-C$\underline{H}_3$; [5] L-Threonine-C$\underline{H}_3$/3-Hydroxybutyrate-C$\underline{H}_3$; [6] vLDL/LDL -(C$\underline{H}_2$)$_n$- [7] L-Lactate-C$\underline{H}_3$; [8] Alanine-C$\underline{H}_3$; [9] TAG–C$\underline{H}_2$CH$_2$-CO; [10] Lysine-C$\underline{H}_2$/Arginine-C$\underline{H}_2$; [11] Arginine γ-C$\underline{H}_2$; [12] Acetate-C$\underline{H}_3$; [13] TAG-CH$_2$C$\underline{H}_2$CH=; [14] Acute-phase glycoprotein molecularly-mobile carbohydrate side-chain N-acetyl-group-C$\underline{H}_3$ functions; [15] Glutamate-β-C$\underline{H}_2$; [16] C$\underline{H}_2$CO / Acetone-C$\underline{H}_3$; [17] Acetoacetate-C$\underline{H}_3$; [18] Glutamate-β-C$\underline{H}_2$ /3-Hydroxybutyrate-C$\underline{H}_2$; [19] Pyruvate-C$\underline{H}_3$; [20] Glutamine-β-C$\underline{H}_2$; [21] Citrate C$\underline{H}_{2A}$/C$\underline{H}_{2B}$; [22] Albumin Lysine/TAG-CH=CH-C$\underline{H}_2$-CH=CH-.

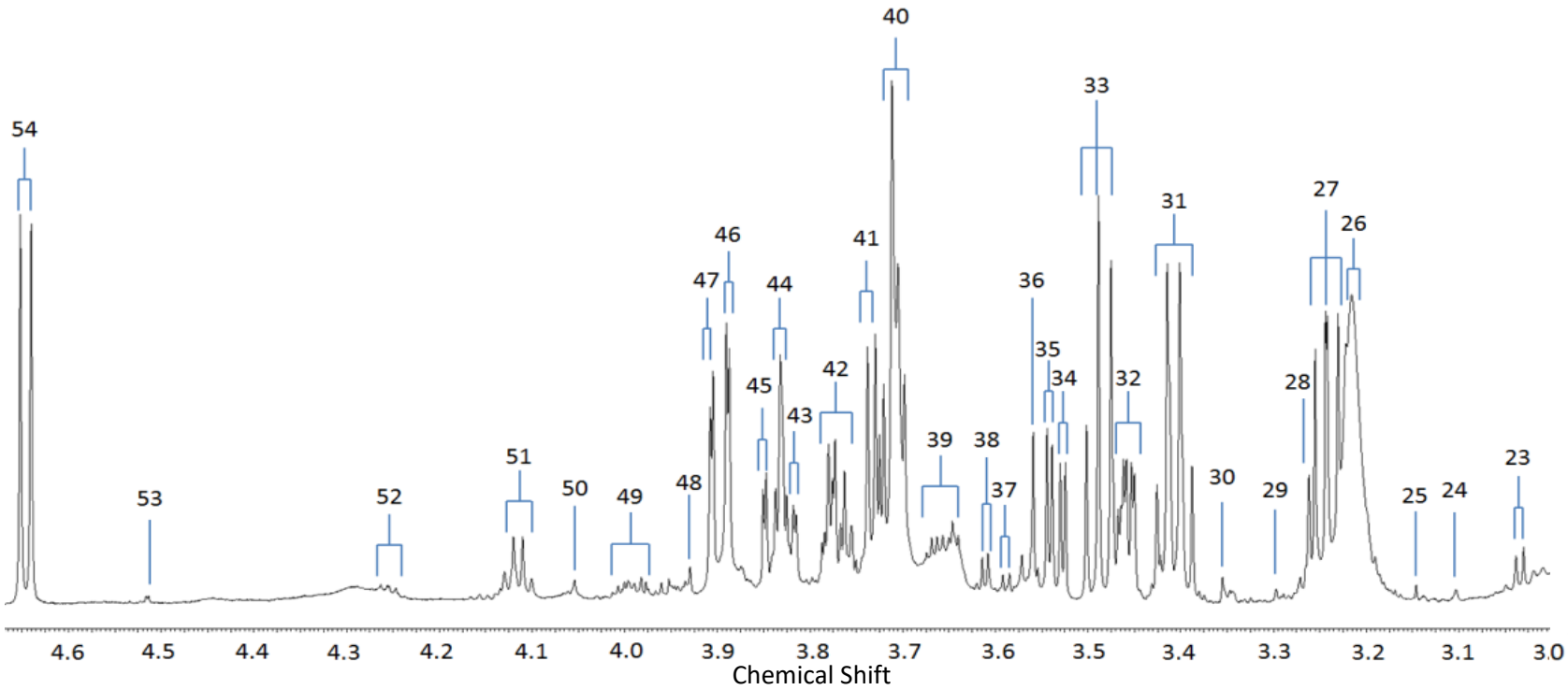


Figure 3.1. Labelled GD plasma sample using CPMG from 3.00 ppm-4.70 ppm; [23] Creatine/Creatinine-C$\underline{H}_3$; [24] unassigned; [25] Dimethylsulphone-C$\underline{H}_3$ [26] HDL-Phospholipid choline headgroup-$N^+(C\underline{H}_3)_3$; [27] D-Glucose-C3-C$\underline{H}$/Taurine-C$\underline{H}_2$; [28] Trimethylamine oxide (TMAO)-C$\underline{H}_3$; [29] unassigned; [30] unassigned; [31] D-Glucose-C5-C$\underline{H}$/Taurine-C$\underline{H}_2$; [32] L-Proline-C$\underline{H}$; [33] D-Glucose-C6-C$\underline{H}$; [34],[35] D-Glucose-C$\underline{H}$; (dd) [36] Glycine-C$\underline{H}_2$; [37] Valine-C$\underline{H}$; [38] Threonine-C$\underline{H}$; [39] Lipid N-C$\underline{H}_2$; [40] D-Glucose-C4/C11-C$\underline{H}$; [41] Galactose-C$\underline{H}$; [42] D-Glucose-C6/C11-C$\underline{H}$; [43] D-Glucose-C11-C$\underline{H}$; [44] D-Glucose-C11-C$\underline{H}$; [45] D-Glucose-C11-C$\underline{H}$; [46] D-Glucose-C11-C$\underline{H}$; [47] D-Glucose-C11-C$\underline{H}$; [48] Creatine-C$\underline{H}_2$; [49] L-Histidine-C7-C$\underline{H}$; [50] Creatinine-C$\underline{H}_2$; [51] L-Lactate-C$\underline{H}$; [52] 3-Hydroxbutyrate-C$\underline{H}$; [53] Dihydroxyacetone-C$\underline{H}_3$ (tentative assignment) ; [54] β-Glucose-C1-C$\underline{H}$.

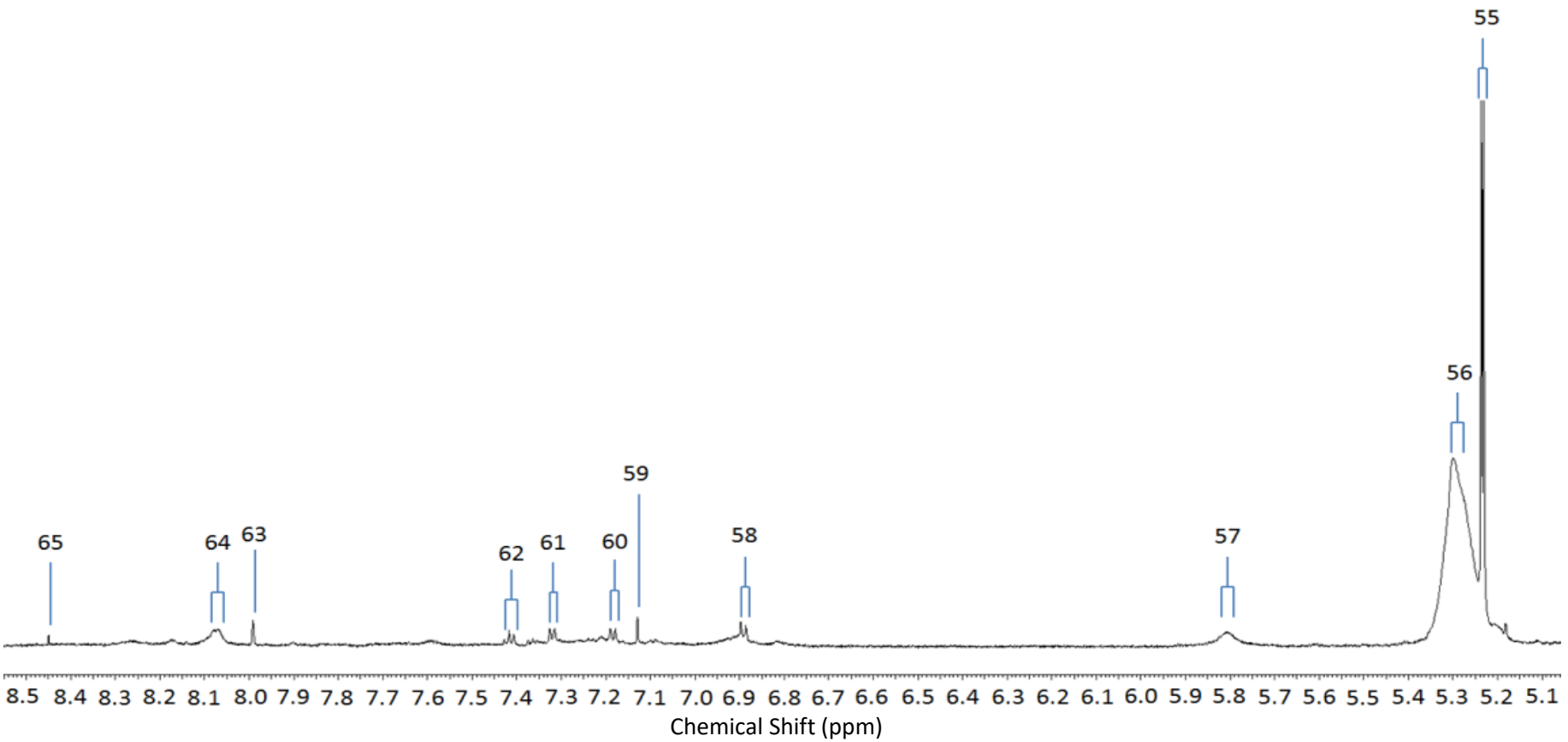


Figure 3.2. Labelled GD plasma sample using CPMG from 5.10 ppm - 8.50 ppm; [55] α-Glucose-C1-C<u>H</u>; [56] TAG-C<u>H</u>=C<u>H</u>-; [57] Urea-CO-N<u>H</u>$_2$; [58] L-Tyrosine-β/ζ-C<u>H</u>; [59] L-Histidine-β-C<u>H</u>; [60] L-Tyrosine-γ/ε-C<u>H</u>; [61] L-Phenylalanine-β/ζ-C<u>H</u>; [62] L-Phenylalanine-γ/δ/ε/-C<u>H</u>; [63] tentative broad protein aromatic residue-C<u>H</u>; [64] tentative broad protein aromatic residue-C<u>H</u>; [65] Formate-C<u>H</u>.

3.5.2 Statistical Analysis of the ^{1}H NMR Profiles of Blood Plasma

Numerous statistical approaches were applied to this dataset. *MetaboAnalyst 4.0,* which provides a number of multivariate statistical outputs (including PCA, PLS-DA, O-PLS-DA and univariate options such as ANOVA, t-test and fold-change analysis). The software also provides clustering and visualisation tools to create dendrograms and heatmaps. This software is also able to classify sample groups by the random forest (RF) and support vector machine (SVM) techniques, and these statistical and machine learning methods have been previously described in Chapter 2.

Data filtering was performed initially in order to remove variables which are unlikely to be of use to modelling data, such as those which contain constant values across the dataset. Normalisation aims to adjust for the dilution factors involved in experiments [101], and in this instance the constant sum-normalisation (CSN) approach was applied. Creatinine normalisation was not performed as it is not used for plasma [101] and, in any case, there are evident differences in fasted and unfasted samples, and thus would therefore not represent a suitable reference substance [102]. TSP was not added to these samples since it can interfere with the matrix by binding to plasma proteins [103] and therefore would be unsuitable as an internal standard for normalisation, unless TSP was added *via* a capillary. Data was then transformed using the cubed root transformation, and scaled by Pareto scaling which is generally recommended for metabolomics data [84]. This method of normalisation will be applied for all of the plasma datasets throughout this work, unless otherwise specified.

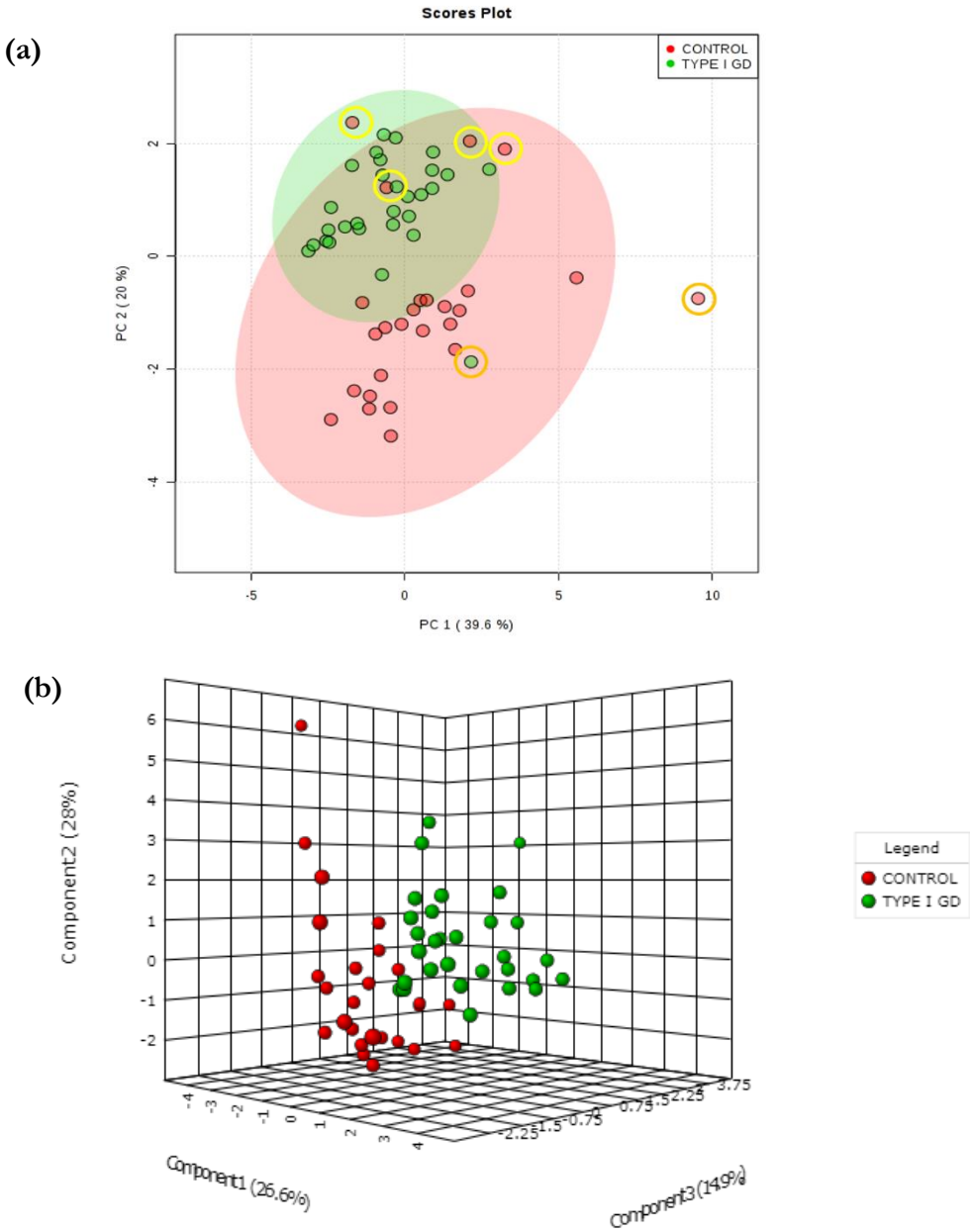


Figure 3.3. (a) PCA scores plot of plasma samples collected from GD1 (green) and fasted and unfasted control participants (red). This diagram shows a plot of PC2 *vs* PC1 accounting for 20 and 39.8% of the dataset variation respectively. 95% confidence ellipses are also shown for both participant classifications. Outliers are highlighted in orange in the control group and yellow circles highlight unfasted control samples. **(b)** PLS-DA scores plot of GD1 plasma (green) *vs* control plasma (red). PC3 *vs* PC2 *vs* PC1 represent 14.9, 28.0 and 26.6% of the total dataset variation respectively. This plot shows some clear discrimination between GD1 and control participants. Cross-validation, utilising a 10-fold leave-one-out cross validation (LOOCV) methodology, gave a Q^2 value of 0.70 on a four-component model. Permutation testing, using 2000 permutations, gave $p < 0.0005$.

PCA is primarily employed to obtain an overview of the degree of separation within the two groups. In addition, PCA can observe clustering between the different classifications in order to detect any outliers in the dataset. The PCA plot of control *vs* GD1 plasma profiles shows

some distinction between the two groups with respect to PC1 and PC2 scores, and revealed no major outliers apart from a single control and a single GD1 group samples (Appendix 2). The former's participant was male, aged 34, with a history of smoking, slightly higher BMI than normal (25.5). The latter profile was from a GD1 participant who was male, 69 years old and on a combination of velaglucerase alfa, hydroxocobalamin, cholecalciferol and calcium carbonate.

The control sample that appears as an outlier highlighted in orange (Figure 3.3 (a)) clearly has lipid disturbances within their plasma ^{1}H NMR spectrum (Appendix 2). 1.20-1.30 ppm was a significant bucket region on the PCA biplot which scored positively for this sample. The outlier attributable to the GD1 participant highlighted in orange arose from an extreme elevation of citrate in the plasma, which is possibly ascribable to a different type of collection tube used (such as one containing a citrate anticoagulant), and so this sample was also removed from the study.

It can be observed that the unfasted control samples coded 501, 502, 503 and 504 are overlapping somewhat with the unfasted GD1 group, instead of the remaining ones in the fasted control group (Figure 3.3, (a)). PCA is an unsupervised methodology, which would suggest that most of the separation observed is due to the fasting status. Interestingly, one sample with GD3 had a different profile to the rest of the control samples in addition to the GD1 profiles, and was clearly distinguishable (data not shown). This suggests that metabolomics analysis with GD3 could provide some interesting metabolic differences. However, with only one sample, it is difficult to establish if this is common across an entire GD3 cohort, and more samples would be required in order to confirm this statistical difference. GD3 is predominantly neurological, and therefore different to GD1 which is predominantly non-neurological.

PLS-DA shows two clear sub-groups for the GD1 participants (Figure 3.3, (b)). Some GD1 participants are within close proximity to the control participants in this analysis, however. Indeed, this could represent disease severity, whereby patients with more severe conditions are further away from the control group and *vice-versa*. Alternatively, this could represent dietary differences within the GD1 cohort. Food diaries were only obtained for the fasted cohort and not the GD1 cohort. Similar statistical tests were performed for the different treatments of GD1, however, and these were insignificant (Appendix 5). Further statistical tests were performed to investigate the impact of participants' age, the results of which were also insignificant and did not account for the sub-clustering noted herein (Appendix 5). PLS-

DA testing demonstrated that the two cohorts were clearly distinguishable (Figure 3.3, (b)). Cross-validation, utilising the 10-fold leave-one-out cross validation (LOOCV) methodology, gave a Q^2 value of 0.70 on a four-component model (Figure 3.3, (b)). Permutation testing, using 2,000 permutations, gave $p < 0.0005$ (Figure 3.3, (b)). These validation tests show that the PLS-DA model is performing efficiently, and that there is a clear difference between the two cohorts in the study. The variable importance in projection plot (VIP) showed from the PLS-DA results were then assigned (Table 3.3). However, it is difficult to ascertain whether the differences herein are ascribable to the fasting and non-fasting of participants, or to the disease pathology. However, considering the VIP information (Table 3.3) and PCA scores plot, it can be considered that the majority of the metabolites are elevated in view of the differential fasting status of the two groups of participants.

δ/ppm	**Elevated or decreased in GD1**	**Assignment (multiplicity)**
0.78 – 0.90	↓	LDL/vLDL-TAG-C<u>H</u>$_3$ (*broad*)
1.91 – 2.01	↓	TAG-CH$_2$C<u>H</u>$_2$CH= (*broad*)/Acetate-C<u>H</u>$_3$ (*s*)
2.32 – 2.35	↑	3-Hydroxybutyrate-C<u>H</u>$_2$ (*m*)
2.97 – 3.02	↓	Albumin lysine residue-ε-N<u>H</u>$_2$ (*t*)
3.16 – 3.22	↓	HDL-Phospholipid choline headgroup-N$^+$(C<u>H</u>$_3$)$_3$
3.29 – 3.31	↑	Glycerol-C<u>H</u>$_2$OH (*m*)
3.62 – 3.64	↑	Lipid Unassigned (*broad*)
3.91 – 3.92	↑	Creatine-C<u>H</u>$_2$ (*s*)/Unassigned (*broad*)
3.93 – 4.00	↑	L-Histidine-C<u>H</u> (*m*)/Phenylalanine-C<u>H</u> (*m*)
4.05 – 4.08	↑	Glyceryl backbone of TAGs-C<u>H</u>$_2$OCOR (*broad*)
4.08 – 4.13	↑	Lactate-C<u>H</u> (*q*)/3-D-Hydroxybuyrate-C<u>H</u> (*m*)
4.22 – 4.26	↑	Threonine-C<u>H</u> /Glycerol backbone of TAGs (*m*)
4.26 – 4.34	↑	Glyceryl backbone of TAGs C<u>H</u>$_2$COR (*broad)*
4.41 – 4.46	↑	Unassigned Lipid (*broad*)
4.62 – 4.65	↑	β-Glucose-C1<u>H</u> (*d*)
7.33 – 7.38	↑	Unassigned-C<u>H</u> (*t*)
7.38 – 7.43	↑	L-Phenylalanine-aromatic-C<u>H</u> (*m*)

Table 3.2. Elevated and depleted metabolites in GD shown by the VIP plot with corresponding bucket regions inclusive of the top 15 most significant metabolites. Abbreviations: GD1, Gaucher's Disease Type 1; TAG, triacylglycerol; HDL, high-density-lipoprotein. Multiplicities are indicated in brackets: multiplet, triplet, doublet and singlet are denoted as (*m*), (*t*), (*d*) and (*s*) respectively.

Glucose, L-phenylalanine and myo-inositol were removed from the PLS-DA analysis since differences between their normalised levels may arise from those observed between fasting and non-fasting participants. These data were then used in subsequent multivariate analysis performed.

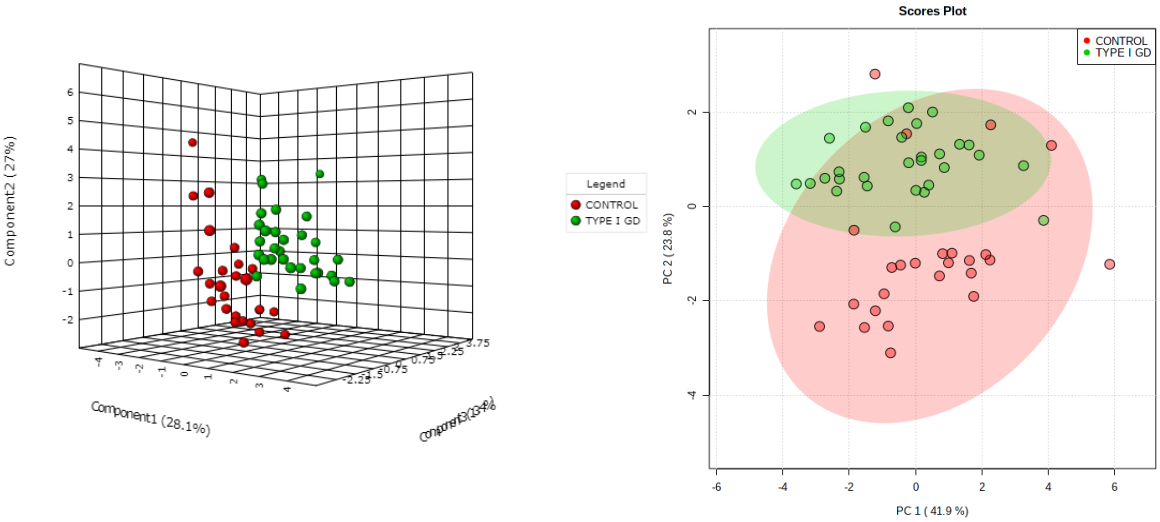


Figure 3.4. (a) PLS-DA scores plot of GD1 plasma (green) and healthy controls (red) with the removal of buckets pertaining to glucose, myoinositol and phenylalanine showing PC3 *vs* PC2 *vs* PC1 axes representing 13.4, 27.0 and 28.1% of the dataset variance respectively. Performance using LOOCV 10-fold methodology gave $Q^2 = 0.66$, and from permutation testing using 2000 permutations, the (*p* value obtained was 0.0005). This plot shows some clear discrimination between GD1 and control participants. **(b)** PCA plot of GD1 plasma (green) and control scores (red). This diagram shows that PC1 and PC2 account for 41.9 and 23.8% of the dataset variation respectively. 95% confidence ellipses are also shown for both participant classifications.

On removal of significant metabolites directly associated with differences observed between fasted and unfasted plasma, a clear separation between these two classifications is always observed, irrespective of the multivariate analysis technique employed for testing. The same four unfasted healthy control samples are scoring/responding in the same manner as the unfasted GD1 participants in the PCA scores plot (Figure 3.4). This effect could be attributable to dietary differences. The PLS-DA plot is able to distinguish between metabolic profile on removal of the significant metabolites found in observed differences between the fasted and unfasted participant profiles such as glucose, myo-inositol and phenylalanine,

suggesting that other metabolites contribute to the differences observed between fasted and unfasted plasma samples.

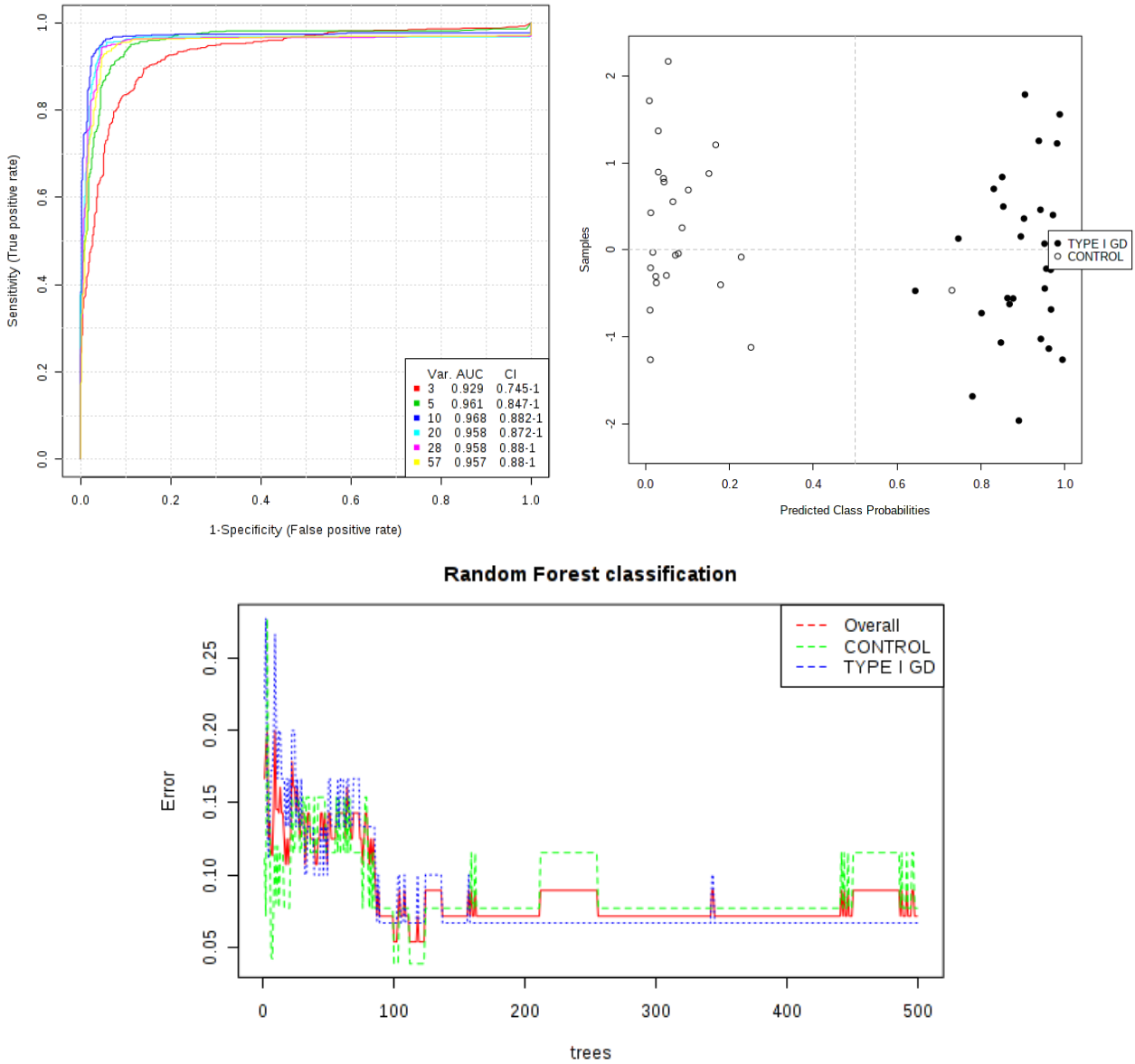


Figure 3.5. (a) Receiver operating curve (ROC) analysis, with the inset showing the number of variables (Var.) in the first column, area under curve (AUROC) in the second column, and 95% confidence intervals (CIs) in the third column. The sensitivity is plotted on the y-axis *vs* selectivity on the x-axis. The highest AUROC value of 0.968 was obtained using 10 variables, each one representing an ISB region. **(b)** Cross-validation matrix showing samples on the y-axis and predicted class probabilities on the x-axis. **(c)** Random forest classification of GD1 plasma *vs* healthy control plasma samples; the green and blue curves representing healthy control and GD1 participants. The red line represents the overall model respectively. The x-axis shows the number of trees built in this model, and the error level (prediction error) is shown on the y-axis. ROC curve analysis can easily distinguish between plasma samples collected from GD1 and control participants using very few variables (Figure 3.5, (a)). The most specific ROC curve test was only using 10 variables which are selected using SVM, which provided an AUROC value of 0.968 and 95% confidence intervals of ±0.0882 (Figure 3.5, (a)). A cross- validation matrix of the ROC Curve shows some cross-over between

participants, since one GD participant is misclassified within in the control group and one control sample is misclassified in the GD1 group (Figure 3.5, (b)). The outlier in this group is one of the unfasted samples, as noted above. This cross- validation gave a 94.6% accuracy with 20 features (Figure 3.5, (b)).

The RF approach gave an out-of-bag (OOB) error value of only 0.0714 using 500 trees and 7 predictors. Out of 26 controls, 2 were incorrectly classified as GD1, and out of a total of 30 of these, 2 were incorrectly classified as controls using this statistical methodology (Figure 3.5, (c)). This may be a reflection of the disease severity, fasting status or high dimensionality of the dataset. Some metabolites may not appear significant in a multivariate model when indeed they are significant, and can be identified in alternative multivariate statistical methodologies instead, or could be found to be statistically significant in univariate models.

Power analysis aids with study design, and ensures that cohort sample sizes are sufficient to detect any metabolic differences observable. Since LSDs are rare, it is important to determine how many samples would be required to provide statistically significant results. Power analysis is not only dependent on sample size but also the magnitude of the effect of interest in the population, along with the statistical significance criterion used in the test. Using a false discovery rate (FDR) setting of 0.1 and a maximum group size of 1,000, power analysis revealed that around 600 samples per group would be required to prove the significant differences found here. However, this is taking into consideration the experimental factors at present which differ *i.e.* diseased samples are non-fasted and the controls are fasted. Less samples may be required if the experimental factors were identical, and therefore it is considered important to repeat this analysis with identical sampling strategies; however, this is challenging in view of the severity of the disease and clinical conditions.

The top 15 significant metabolites in each multivariate statistical test performed are listed in Table 3.3. Interestingly, some methodologies consider the same metabolites statistically significant, for example the 4.26 – 4.34 ppm bucket region, and other metabolites are only statistically significant for a single bucket variable, for example the 2.20 – 2.25 ppm chemical shift region (Table 3.3).

δ/ppm	Elevated or decreased in GD	Assignment (multiplicity)	Statistical Test		
			PLS-DA VIP value	ROC Curve	RF
4.41 – 4.46	↑	Unassigned (*broad*)	2.71	✓	
4.26 – 4.34	↑	TAG glycerol backbone (*broad*)	2.50	✓	✓
4.08 – 4.13	↑	Lactate-C<u>H</u> (*q*) / 3-D-Hydroxybuyrate-C<u>H</u> (*m*)	2.03		✓
4.22 – 4.26	↑	Threonine-C<u>H</u> /TAG-glycerol backbone (*m*)	1.77	✓	✓
0.78 – 0.90	↓	TAG-LDL/vLDL-terminal-C<u>H</u>$_3$ (*broad*)	1.55	✓	✓
1.91 – 2.01	↓	TAG-CH$_2$C<u>H</u>$_2$CH= (*broad*) / Acetate-C<u>H</u>$_3$ (*s*)	1.54		✓
3.62 – 3.64	↑	Unassigned (*broad*)	1.40	✓	✓
2.32 – 2.35	↑	Glutamate-β--C<u>H</u>$_2$ (*m*)	1.35	✓	✓
2.76 – 2.80	↑	TAG unsaturated fatty acids (*broad*) - C=CH-C<u>H</u>$_2$-CH=C	1.34	✓	✓
7.16 – 7.19	↑	L-Tyrosine-γ/ε-C<u>H</u> (*s*)	1.20		
1.20 – 1.30	↓	vLDL/LDL-bulk-chain-(CH$_2$)$_n$- (*broad*)	1.18		
4.03 – 4.05	↑	Creatinine-C<u>H</u>$_2$ (*s*)	1.17	✓	✓
2.97 – 3.02	↓	Unassigned (*m*)	1.13	✓	✓
7.79 – 8.01	↑	Histidine-C<u>H</u> (*s*)	1.13		
0.94 – 0.96	↑	Leucine-C<u>H</u>$_3$ (*d*)	1.11		
2.35 – 2.37	↑	Pyruvate-C<u>H</u>$_2$ (*s*)	1.09	✓	✓
1.45 – 1.48	↓	Alanine-C<u>H</u>$_3$ (*d*)	1.03	✓	✓
2.90 – 2.92	↑	TMAO/Unassigned (*broad*)		✓	
2.63 – 2.69	↓	C=CC<u>H</u>$_2$C=C (*broad*)/Citrate CH$_{2A}$/CH$_{2B}$ (*d*)		✓	✓
3.33 – 3.35	↑	Unassigned (*broad*)		✓	
2.20 – 2.25	↑	Acetone-C<u>H</u>$_3$ (*s*) / C<u>H</u>$_2$CO (*broad*)			✓

Table 3.3. Top 15 statistically significant buckets, showing up- or downregulations in GD1, and for which statistical tools show that these metabolites are significant. VIP Scores are reported in the PLS-DA column. Abbreviations: GD1 Gaucher's disease, type 1; HDL, High-density-lipoprotein; LDL, Low-density-lipoprotein; TAG, Triacylglycerol; VIP, Variable Importance Parameter.

3.5.3 Univariate Statistics Performed on Plasma Profiles

^{1}H NMR profile frequency buckets which were found to be significant in univariate testing were then assessed using box plots and ANOVA (Appendix 3). The PCA loadings plot identified elevations in 4.08 – 4.13 (-0.06, 0.24), 4.26 – 4.34 (-0.12, 0.13), 4.41 – 4.46 (-0.16, 0.09), and 4.22 – 4.26 (-0.07, 0.09) ppm bucket regions (PC1 and PC2 loadings are shown respectively in brackets), from observation of the extreme loadings on PC1 and PC2. These are also displayed as box and whisker plots (Appendix 3). These bucket regions are attributable to lactate-C<u>H</u> (*q*)/3-D-hydroxybuyrate-C<u>H</u> (*m*), TAG glycerol backbone (*broad*), unassigned broad singlet (*s*), and threonine-CH (*m*). The ROC curve analysis identified bucket regions attributable to TAG glycerol backbone, TAG-LDL/vLDL-terminal-C<u>H</u>$_3$, TAG unsaturated fatty acids and TAG-CH$_2$C<u>H</u>$_2$CH= (*broad*)/Acetate-CH$_3$ (*s*) to be significant bucket regions, with the minimum AUROC value for these particular predictor variables to be 0.823 (Appendix 3). Both univariate test results agree with the above multivariate statistical analyses which were performed.

3.5.4 Enrichment Analysis on Plasma Dataset

Enrichment analysis considers disturbed metabolites in the GD1 plasma dataset, and compares it to that of the healthy controls. Such differences are then explored with special reference to these datasets that are defined by disease or biological locations. This helps to identify correlations with multiple metabolites that might be disturbed in a disease or location. Metabolite set libraries used include the location-based set including organs, tissue and sub-cellular locations, and the disease associated metabolite set (Blood) containing 344 reported metabolite sets in blood.

Location-Associated Enrichment Analysis

Enrichment analysis using *Metaboanalyst 4.0* revealed skeletal muscle as the most significant organ impacted in the metabolite dataset (FDR corrected *p*-value 2.76 x 10^{-8}) in view of an upregulation in creatine/creatinine, glutamine, arginine, leucine and pyruvate in GD1 disease. The skeletal muscle manifestations in GD1 are well documented in the literature [104]–[107]. Moreover, the second most significant location was revealed to be the muscle (FDR-corrected *p* value 3.49 x 10^{-7}). The third and fourth most significant organs were the spleen (FDR-corrected *p* value 6.01 x 10^{-7}) and bladder (FDR-corrected *p* value 1.95 x 10^{-6}). Significance of the impact of splenomegaly in GD1 is also frequently reported [104], [107]–[109]. Muscle is a less well documented manifestation of GD, but more recently, correlations have been established between GD1 and myopathy, although this has been variable amongst patients with this disorder [110]. All three of these organs are features that are affected by GD1 disease pathology.

Other locations that were identified as significant included the bladder, epidermis, placenta, intestine, peroxisome, pancreas and prostate. It is unsurprising that other organs are significant, in view of the age of the GD1 patients, and the number of medications that the patients are taking, which suggests that there are other age-related metabolic changes and possibly age-related diseases. The liver was also significant, but to a lesser extent, and this organ often inflamed in GD1 [104]. It would be unsurprising to observe the liver as a significant organ, due to the number of drugs the patients were taking. Nevertheless, GD is a heterogeneous multi-symptomatic disorder, hence metabolic disturbances are variable from patient to patient. Additionally, due to the other medications the participants are taking, it cannot be omitted that some of the other organs are impacted because of pharmacological related changes.

Disease-associated Enrichment Analysis

Critical illness, early markers of myocardial injury, cirrhosis, rhabdomyolysis, diabetes, N-acetyl glutamate deficiency, heart failure and inflammatory diseases were all significant in the disease associated enrichment analysis performed (FDR corrected *p*-value < 0.01181). Cirrhosis metabolomics profiles correlating with the study herein is unsurprising due to hepatomegaly observed in GD1 patients [111]. Furthermore, rhabdomyolysis is a condition related to muscle injury which also correlates with the disease pathology that observed in the form of myopathy in GD1 patients [110].

Metabolites affected by gender (FDR corrected *p*-value = 2.88 x 10^{-7}) and age-related (FDR corrected *p*-value 2.88 x 10^{-7}) metabolites were two of the most significant metabolite sets in enrichment analysis. The participants are predominantly sex- and age-matched, and hence this finding was unexpected.

3.5.5 ^{1}H NMR Urinary Profile Results

^{1}H NMR profiles acquired were referenced to TSP, overlaid and labelled using literature sources [60], [112] and reference to the HMDB. All the GD patients involved in the study were on a combination of medications, which may have an effect on NMR spectra in two different manners. Drugs and their metabolites may produce resonances which overlap with metabolite ones, and therefore need to be removed in order to not appear as an anomaly following multivariate/univariate statistical analysis. In addition, metabolomics studies which considered the drugs and supplements taken by the patients were monitored, so changes in endogenous biomolecule metabolism arising from drug interactions or other xenobiotics could be ascertained in depth, together with the impact this may have on the results acquired. A literature review performed herein established whether the drugs and their metabolites were typically excreted in urine, and if so, their corresponding urinary concentrations. Additionally, it was observed how, if at all, drugs impacted on the metabolomic profiles observed, and how this would impact on the study performed here. Drugs and their metabolites were removed from ^{1}H NMR spectra when detectable by removal of the bucket regions containing such xenobiotic resonances.

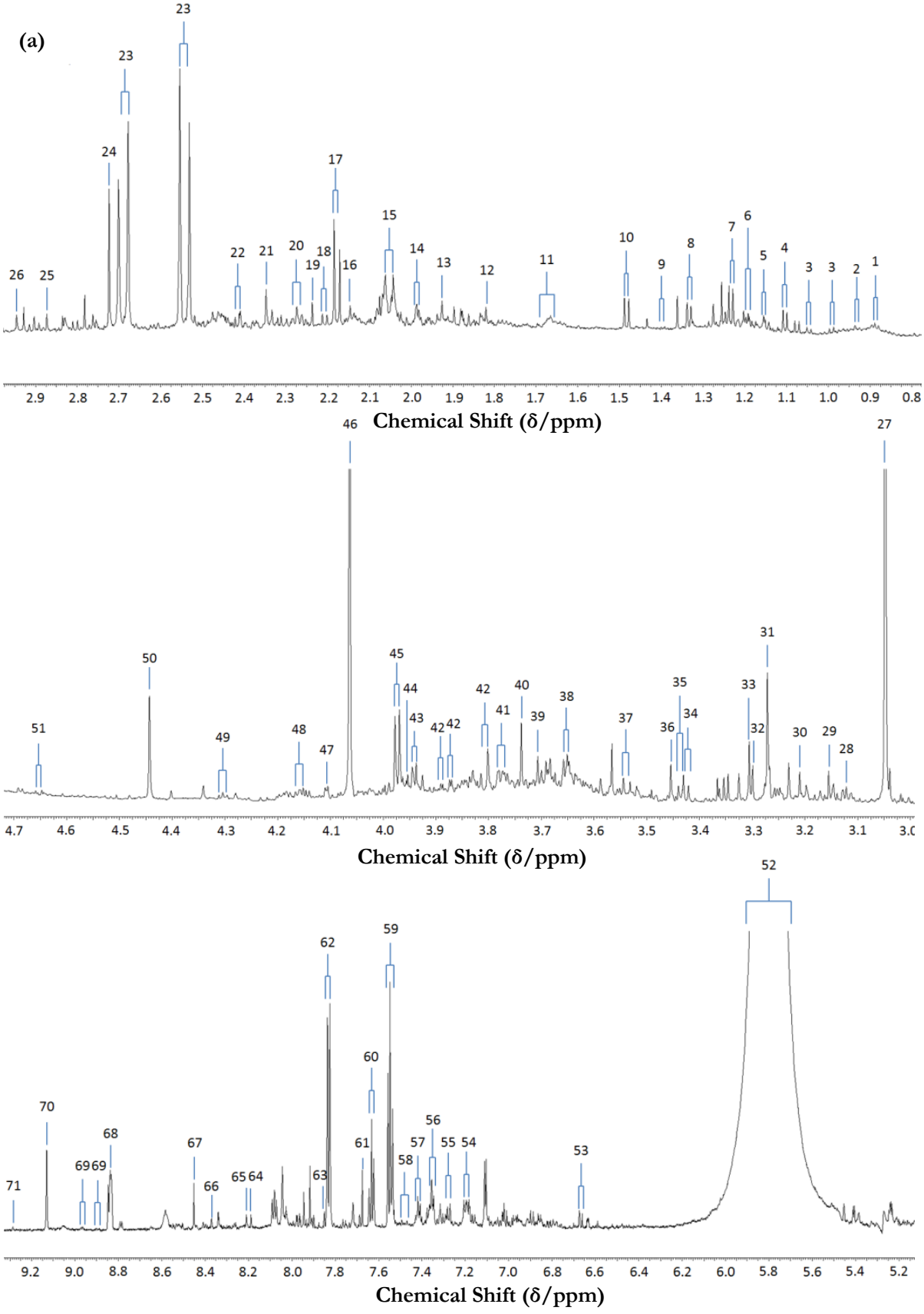


Figure 3.6. (a) 700 MHz ^{1}H NMR spectra of typical GD1 and control urine samples using the NOESY PRESAT pulse sequence from 0.35-0.75 ppm, highlighting the region where

bile acids C18-methyl function resonances appear. **(b)** Corresponding spectrum of a typical GD1 urine sample using NOESY PRESAT from 0.85-9.20 ppm. Abbreviations: [1] Isoleucine-C$\underline{H}_3$; [2] Leucine-C$\underline{H}_3$; [3] Valine-C$\underline{H}_3$; [4] Methylsuccinate-C$\underline{H}_3$; [5] 2-Oxoisovalerate-C$\underline{H}_3$; [6] Amino-isobutyrate-C$\underline{H}_3$; [7] 3-D-hydroxybutyrate-C$\underline{H}_3$; [8] Lactate-C$\underline{H}_3$; [9] Lysine-γ-C$\underline{H}_2$; [10] Alanine-C$\underline{H}_3$; [11] Adipate-C3-C$\underline{H}_2$; [12] Amino adipate-C$\underline{H}_3$; [13] Acetate-C$\underline{H}_3$; [14] Lysine-β-C$\underline{H}_2$; [15] N-Acetyl-X-C$\underline{H}_2$; [16] Glutamine-γ-C$\underline{H}_2$; [17] Acetone-C$\underline{H}_3$/Unassigned; [18] Unassigned; [19] Acetoacetate-C$\underline{H}_3$; [20] Hydroxygluterate-δ-C$\underline{H}_2$; [21] Pyruvate-C$\underline{H}_3$; [22] Glutamate-γ-C$\underline{H}_2$; [23] Citrate-C$\underline{H}_{2A}$/C$\underline{H}_{2B}$; [24] Dimethylamine-C$\underline{H}_3$; [25] Trimethylamine-C$\underline{H}_3$; [26] Dimethylglycine-C$\underline{H}_3$; [27] Creatine/Creatinine-C$\underline{H}_3$; [28] Malonate-C$\underline{H}_3$; [29] L-Histidine-C$\underline{H}_2$; [30] D-Glucose; [31] Trimethylamine N-oxide(TMAO)-C$\underline{H}_3$; [32] Taurine-C$\underline{H}_2$-N$\underline{H}_3$+; [33] Trimethylamine-C$\underline{H}_3$; [34] D-Glucose-C$\underline{H}$; [35] Cystine-γ/ζ-C$\underline{H}$; [36] Taurine-C$\underline{H}_2$-SO_3; [37] 3-Hydroxyphenylacetate-C$\underline{H}_2$; [38] Mannitol-γ-C$\underline{H}$; [39] Phenylacetylglutamine-C$\underline{H}_2$; [40] 4-Methyloxyphenylacetate-C$\underline{H}_2$; [41] Lysine-β-C$\underline{H}$; [42] Mannitol-C$\underline{H}_2$; [43] Creatine-C$\underline{H}_2$; [44] Glycolate-C$\underline{H}_2$; [45] Hippurate-C$\underline{H}_2$; [46] Creatine-C$\underline{H}_2$; [47] Pyroglutamate-C$\underline{H}$; [48] Lactate-C$\underline{H}$; [49] Tartrate-C$\underline{H}$; [50] Dihydroxyacetone-C$\underline{H}_3$; [51] β-Glucose-C$\underline{H}$; [52] Urea-CO-N$\underline{H}_2$; [53] 2-Furoylglycine-γ-C$\underline{H}$; [54] Indoxylsulphate-C6/C9-C$\underline{H}$; [55] Indoxylsulphate-C2-C$\underline{H}$ / Phenylalanine-C2/C6-C$\underline{H}$; [56] Phenylalanine-C3/C4/C5-C$\underline{H}$; [57] Indoxylsulphate-C2-C$\underline{H}$; [58] Indoxylsulphate-C$\underline{H}$ / Benzoate-C$\underline{H}$; [59] Hippurate-C3/C5-C$\underline{H}$; [60] Hippurate-C3/C5-C$\underline{H}$; [61] 1-Methyl-Histidine-C$\underline{H}$; [62] Hippurate-C2/C6-C$\underline{H}$; [63] Histidine-β-C$\underline{H}$; [64] Hypoxathine-C2/C7-C$\underline{H}$; [65] 1-Methylnicotinamide-α-C$\underline{H}$; [66] Quinolinate-C6-C$\underline{H}$; [67] Formate-C$\underline{H}$; [68] Trigonelline-C4/C6-C$\underline{H}$; [69] 1-Methylnicotinamide-C$\underline{H}$; [70] Trigonelline-β-C$\underline{H}$; [71] N-Methylnicotinamide-C$\underline{H}$.

Not all assignments could be finalised in view of the complexity of urine as a biological sample, since there are over 200 ^{1}H NMR-detectable biomolecules therein [60], in addition to overlap problems sometimes encountered with signals arising from drug resonances and their corresponding metabolites. Further analysis using 2D NMR techniques would be required to confirm all resonance assignments.

Acetaminophen and it's metabolites (Appendix 4) are excreted in human urine, and this is already extensively established in the metabolomics literature [113], [114]. These are examples of signals which are required to be removed prior to metabolomics analysis, or alternatively the samples collected from participants receiving this particular medication can be removed and hence excluded from the analysis. In view of the number of drugs the participants are receiving, all of the samples were kept in the model and the metabolite resonances that arose from drugs and their metabolites were removed. If fewer participants were on medication, these participants could be removed, and the study could be performed with a smaller number of participants. Additionally, an aspirin metabolite, salicyluric acid, was also identified in human urine (Appendix 4), and this has been confirmed by reference to the HMDB and further literature [113]. Interestingly, although a patient was receiving ibuprofen, resonances attributable to this drug that have previously been reported in the

literature were not observed [115]. The half-life of ibuprofen is ~2 hours and this may be the reason that the parent compound or its metabolite resonances were not observed, i.e. it is rapidly excreted in the urine. Although the medications of all individuals participating in the study were obtained, the time of administration of the drugs was not provided, information which would help in this metabolomics investigation if it was required to investigate drug metabolism further.

3.5.6 Statistical Analysis of Urinary ^{1}H NMR Profiles

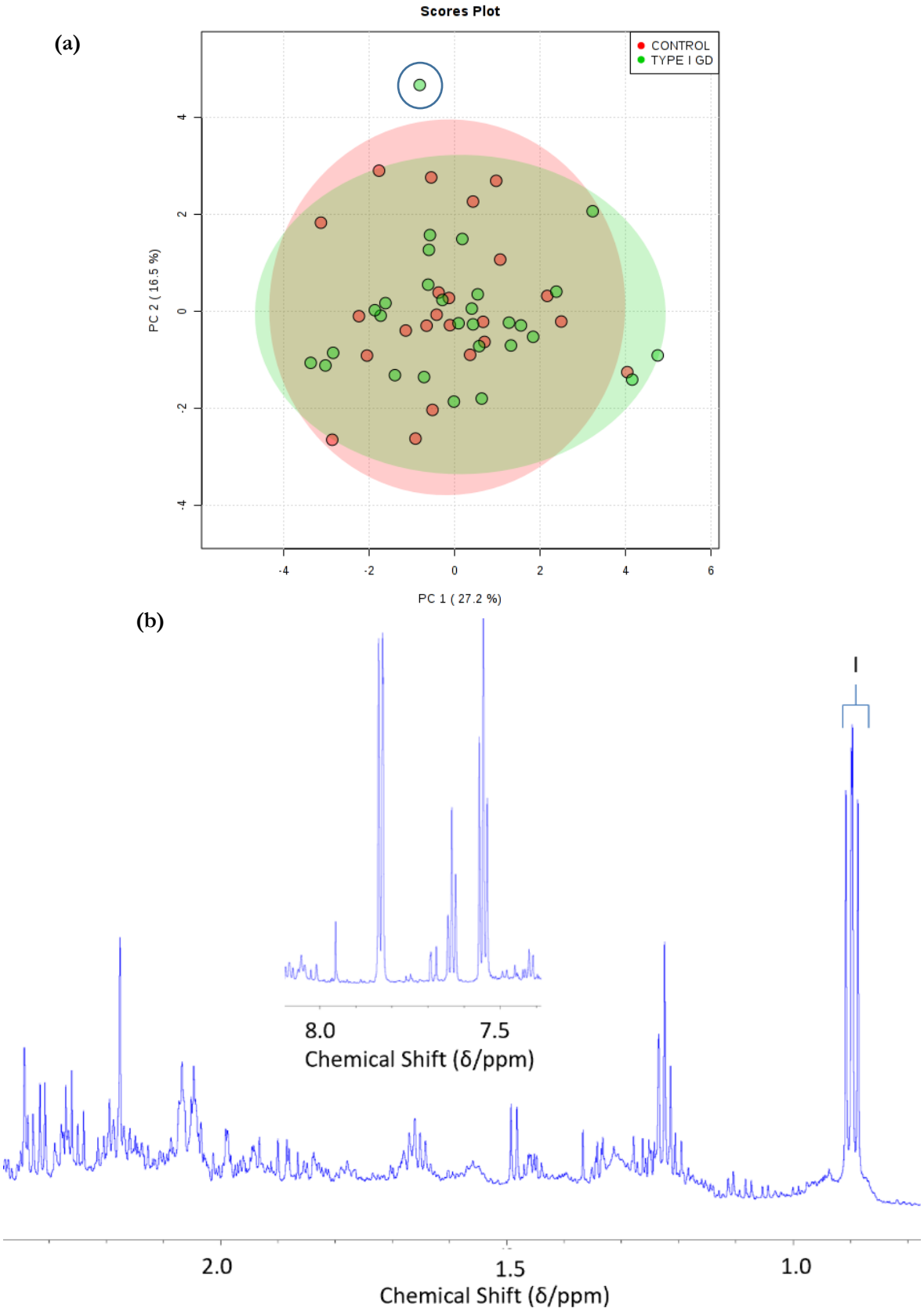


Figure 3.7. (a) PCA scores plot of control (red) *vs* type I GD (green) urinary profiles showing that PC1 and PC2 represented 27.2 and 16.5% dataset variance respectively. 95% confidence ellipsoids are also shown for both participant classifications. An 'outlier' sample from a patient receiving eliglustat therapy is circled in blue. **(b)** Potential eliglustat resonance in ^{1}H NMR profile labelled [I] which corresponds to its terminal-CH_3 function, but this could also be attributable to pregabalin's-CH_3 resonance. Aromatic hippurate signals are shown on the inset image with resonances located at δ =7.54, 7.56 and 7.82 ppm.

PCA analysis shows no clear distinction between controls and GD1 (Figure 3.7). There is one outlier in the data in the GD1 disease group outside the confidence ellipses which is the only patient taking eliglustat. Eliglustat is primarily excreted in the faeces in view of poor oral absorption, but has also been observed in the urine to a lesser extent and has been described as having unharmful oxidised metabolites [116], [117]. On further observation of the scores for the PCA plot and the corresponding spectrum a large resonance at δ = 0.899 ppm that was loading positively on PC2. A large elevation in hippurate was also observed, which could be a response to the eliglustat, since this is the highest level of hippurate found in any of the ^{1}H NMR spectra acquired here. These aromatic signal bucket regions and the doublet at δ = 3.95 ppm were also scoring positively on PC2. Further confirmation using 2D NMR would be required in order to confirm the rest of the drug metabolites signals. This particular participant was also receiving pregabalin therapy, and this drug also has a methyl group with a similar chemical shift value to that of eliglustat at δ = 0.90 ppm [118] and therefore it is proposed that this signal is from either one of these drugs.

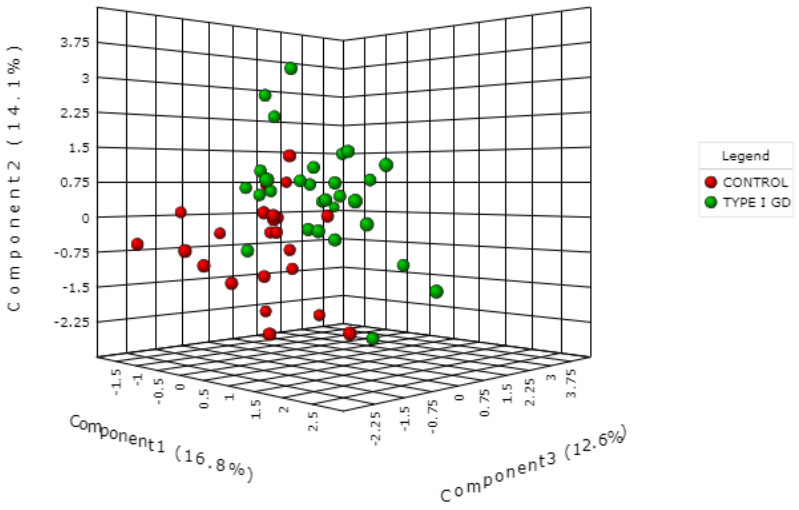


Figure 3.8. PLS-DA scores plot of control *vs* Type I GD showing a plot of PC3 vs. PC2 vs. PC1 representing 12.6, 14.1 and 16,8% of dataset variance respectively. 10-fold cross-validation was performed, and a model containing 3 components gave a Q^2 =value of -0.07. However, permutation testing was also performed, and this gave a *p*-value of 0.068 (not significant) with 2,000 permutations using separation by distance.

PLS-DA did not distinguish between control and GD1 urinary profiles. It appears that some control participants are loading in a similar manner to that of GD1 participants (Figure 3.8). Regardless, the model appears to be performing poorly and the cross validation of the data showed poor accuracy and precision showing that the model was overfitting. Furthermore, the permutation test showed that the model was insignificant.

SVM was applied to this dataset and it was found to provide a 51.16% error rate with five variables, this was the lowest error rate which shows that SVM is not suitable for trying to classify between these two groups. A range of different validations were used including bootstrap, 10-fold cross validation and LOOCV. k-Means clustering and self-organising maps were also unable to distinguish between the groups.

Univariate analysis was also completed, and this found no statistically significant metabolites in all the tests performed, including fold-change analysis, t-tests and volcano plots.

Multivariate Modelling of Gaucher Disease Treatments

Multivariate modelling of the two different treatment types were explored in order to distinguish if they had any significant impact on metabolic patterns in urine.

PCA analysis showed a clear outlier in the initial analysis. This may be because this participant is the only individual taking omeprazole and there are clear resonances that could be attributable to this drug. However, full identification would be required by 2D NMR strategies. The PCA analysis shows no real separation in the group (Appendix 5). This outlier was removed before further supervised analysis.

A PLS-DA scores plot showed no significant difference between metabolites in the two different treatment types employed herein, since there was no separation observed between confidence ellipses (Appendix 5). The PLS-DA analysis revealed low levels of accuracy and precision. Additionally, cross-validation provided a p-value > 0.05. Upon permutation, the test could not distinguish accurately between the two groups. Similarly, sPLS-DA and O-PLS-DA approaches confirmed no distinction between the two groups.

With the use of SVM analysis and using 37 variables, the lowest error rate was 45.8% using all three combinations of validation methods including LOOCV, bootstrap and 10-fold cross validation. This was the lowest percentage error reported amongst all the variables. RF could not distinguish between sample groups either, with OOB error at 0.538 using 1,000 trees and 7 predictor variables.

Fold change, t-test and volcano plot revealed no significant differences in metabolites in urinary profiles on comparison of the two different treatments.

Multivariate Modelling of the Impact of Age in GD participants

Additional analysis was performed to explore if age groups could be distinguished in the GD1 patient cohort. Disease severity index scores were not provided as part of the dataset, and therefore age was considered as one of the variables that could help indicate the metabolic age and severity of the disease in the participants. The average age of the cohort was 50 years, so the cohort was split into groups of below and above 50 years.

No significant differences were noted in GD participants using PCA or PLS-DA. The PLS-DA model did not find a significant difference between the two groups in view of the poor Q^2, R^2 and accuracy scores reported. There appears to be one main outlier in the group in the PCA plot and this is the participant taking eliglustat, rather than the other two receiving enzyme medications for GD treatment. It was found that no significant differences were found in terms of age of participants using computational intelligence techniques such as SVM, which provided a 58.7% error with 9 variables and this was with all forms of validation testing. In univariate testing, one metabolite was significant when using fold-change testing, and it was a trigonelline bucket region (δ = 9.10-9.15 ppm); this is a metabolite of caffeine, which is often found to be elevated in coffee drinkers [119]. k-Means clustering was unable to distinguish between the different age groups, but interestingly the three clusters that were formed included the eliglustat participant on its own.

3.6 Discussion

3.6.1 Potential Plasma Biomarkers

Metabolites linked to GD pathology

Lipoproteins are responsible for the transportation of lipids in the circulatory system [120]. HDL removes lipids from cells, and therefore high levels of HDL are associated with decreased risk of cardiovascular events [120]. Conversely, high levels of LDL are commonly associated with an increased risk of such events [120]. Interestingly, buckets relating to HDL-, LDL- and vLDL-TAGs are all decreased in the GD1 plasma profiles, despite the fasting period involved for the healthy control participants, since all these acylglycerols can

be obtained from the diet and GD participants were not fasted, unlike the control ones. This is the first time these lipid disturbances have been observed *via* ^{1}H NMR analysis. These lipoproteins are responsible for transporting cholesterol. Low levels of HDL cholesterol reverse cholesterol removal, and this may lead to disturbances in cholesterol transport and its deposition; this biomolecule is one of the major contributors of gallstones in GD1 patients [121].

Unlike LDL and vLDL cholesterol, HDL cholesterol has been shown to decrease after eating, and therefore the depletion may be exaggerated here in this study due to the GD1 patients being non-fasted [122]. Nevertheless, all these markers have been previously recognised as depleted in GD1 patients in view of metabolic changes [123], [124], and serve as evident lipid disturbances in the GD1 plasma profiles acquired. On average, GD1 patients are older than the control participants within this study (Tables 3.0 and 3.1), which would predispose to higher levels of LDL cholesterol in this group [125]; However, this is not the case here. Indeed, the inverse effect was observed for HDL cholesterol, which diminishes with age, and therefore the results acquired here could be ascribable to the slight difference in age ranges between GD1 and control participants, since the age-matching of the two groups for comparison is imperfect [125]. Additionally, multivariate modelling was performed in order to determine if metabolic differences were distinguishable by age, but this effect was not found to be significant.

Low HDL cholesterol is associated with an increased risk of cardiovascular problems, whilst increased HDL is associated with longevity of life [125], [126]. However, a previous study on GD participants showed that low HDL cholesterol does not correlate with atherosclerosis assessed by artery intima-media measurements [127]. Low levels of HDL cholesterol are already recognised as a marker to diagnose patients of Ashkenazi Jewish ancestry with GD1; however, confirmatory testing of GD1 is almost always performed with leukocyte glucocerebrosidase assays [128].

LDL cholesterol was also found to be low in GD participants who are on average slightly older than the control participants. This is interesting, since typically LDL cholesterol should be higher in older participants and so it could be expected LDL cholesterol to be elevated in GD1 as a result of non-fasting [122]. Indeed, vLDL cholesterol was also found to be low in GD1 patients.

It has been previously hypothesised why these disturbances occur in GD1 patients, with potential causes being ascribable to: (i) malnutrition; (ii) displacement of cholesterol with

glucosyl ceramide; and (iii) reduced synthesis in the liver [123]. The most likely cause of disturbances in LDL/HDL cholesterol levels are believed to be attributable to a reduction in the level of the apolipoprotein B100 and apolipoprotein A1, and an elevation in apolipoprotein E in the liver, which are responsible for LDL and HDL biosynthesis [124]. Alternatively, this could be attributable to a lack of ABCA1, similarly to NPC, or LCAT proteins which control cholesterol trafficking [127]. Both NPC and GD exhibit liver inflammation and a disrupted cholesterol transport [26], which could give rise to these depleted plasma markers.

There has previously been described the relationship between gallstones and low HDL/LDL cholesterol in GD1 patients, gallstones were correlated with low HDL/LDL cholesterol and elevated severity score index, also they were found to be composed mostly of cholesterol [121]. Previous studies have suggested these disturbances are also present in β-glucocerebrosidase (GBA) carriers [129], and have suggested that these markers are inversely correlated with disease severity [123]. Upon treatment with ERT, patients' HDL cholesterol levels are elevated, but do not normalise, and LDL cholesterol levels remain unchanged [130]. Although this preliminary study has not included carriers, this may be an interesting addition to see whether metabolomics could distinguish between all three groups. However, this is the first time that these disturbances have been detected using an ^{1}H NMR metabolomics approach in conjunction with other potential biomarkers.

Chronic liver disease and cirrhosis were found to be in the top 5 leading causes of death pertaining to GD1 patients [131]. Elevated lactate has been previously described as a biomarker for acute liver failure [132], which would corroborate with disease profiles here as GD1 participants have hepatomegaly and an elevation in plasma lactate levels. Hypoxia and cirrhosis are also highly correlated to elevations in lactate. Furthermore, necrosis is common in GD1 patients, which is also highly correlated to elevated lactate levels in plasma, and is more prominent in patients that have not been treated with ERT within 2 years of diagnosis [133]. Decreases in HDL and elevations in lactate together, formulate the hypothesis that there is a shift from aerobic glycolysis to anaerobic and lipid metabolism, which is also observed in the blood serum of patients with acute-on-liver failure [134]. Elevations in lactate and creatinine have also been previously correlated with cirrhosis [134], [135], and both are upregulated in the GD1 profiles observed in this study. Moreover, lactate has also been observed to be elevated in plasma of malnourished children [136] and patients with anaemia, sepsis and trauma [137]. Malignancy has also been found to be the top cause of death in GD1 patients [131], and this is correlated with elevated lactate levels [137]. Blood lactate is a

biomarker that has been recognised for mortality, and upregulations of this marker is inversely correlated with survival, and has been shown in numerous studies [138]. Overall, lactate is a non-specific marker and it could be linked to numerous disorders in GD1.

Tyrosine and phenylalanine were also upregulated in GD1 patients which have been previously found to serve as early markers of liver injury [134], [135]. In addition, acetone, which is a ketone body, was also increased in GD1 plasma profiles (Table 3.3), which also indicates liver injury, which would fit with GD1 pathology. However, upregulations of acetone in plasma have also been correlated with many other conditions, such as uncontrolled diabetes [139], chronic renal insufficiency [140] and pancreatic cancer [141], for example. Amino acid disturbances have been found in patients with chronic liver disease, including upregulations of both tyrosine and phenylalanine [142].

Formate plays a key role in disrupting the mitochondrial electron transport chain and energy production by the inhibition of cytochrome oxidase activity [99]. This is the terminal acceptor of the electron transport chain [99]. Consequently, cell death can occur in view of depletion of adenosine triphosphate (ATP), which reduces the energy requirements of cells, and hence functions cannot be maintained [99]. This could support the hypothesis that mitochondrial dysfunction is occurring in GD1 [143]. The elevation of formate in the blood plasma is showing the bodies response attempting to remove the excess formate from within the cells. Again, this could be another key biomarker for GD.

Histidine is a metabolite which suppresses inflammation [99]. This metabolite is observed to be elevated in GD1 blood plasma (δ = 3.93 – 4.00 ppm) using multivariate analysis (Table 3.3). However, phenylalanine also resonates within this bucket region. Phenylalanine, as noted above, is considered as having little clinical significance with the exception of elevations being observed in patients with chronic liver disease.

3-D-hydroxybutyrate (3-HB) increases with fasting, exercise [144], ketogenic diets and calorie restriction [145]. 3-HB has also been demonstrated to be elevated after eating [122]. Thus, the increase in unfasted GD1 patient's plasma may be significant. 3-HB is often used as a source of energy when blood glucose levels are depleted; however, this ketone body was found to increase despite the blood glucose levels in participants being significantly elevated.

Moreover, the HDL-phospholipid choline headgroup resonance is significantly decreased in GD1 patients. Choline is synthesised in the body of humans, but can also be obtained from dietary sources. This effect was observed despite the fasting status of GD1 patients. Choline is an important molecule for liver function and skeletal muscle [146], and a depletion in

choline can cause liver damage [147]. Skeletal muscle contains large amounts of choline and a small amount of acetylcholine, its precursor. Both the liver and bone marrow dysfunction are involved in GD1. This could show viability as a potential biomarker for GD1.

A bucket attributable to leucine (δ = 0.94 – 0.96 ppm) is elevated in GD participants. This bucket is slightly overlaid by the lipid resonance, but this resonance without inclusion of the leucine signal is observed to be reduced in GD rather than increased, which contrasts with the lipid signal. Leucine is an essential branched-chain amino acid, and breaks down in a ketogenic manner producing acetoacetate and acetyl-CoA [99]. Leucine was shown to be elevated in GD participants and this could be linked to changes in the mitochondria and possibly links to other metabolic changes described herein. Alternatively, leucine can be obtained from the diet and this may be the reason why it is elevated in GD1 patients.

Alanine is downregulated in GD profiles on observation of the bucket region 1.46 – 1.48 ppm. Alanine is a non-essential amino acid and is synthesised by the body, and can be acquired from DNA breakdown or conversion of pyruvate [148]. Interestingly, alanine decreases on fasting, and GD1 participants are unfasted hence this may be a marker of interest for this disease. Additionally, alanine can be obtained from the diet, so an elevation in alanine in GD participants is expected, and not a decline. Depletion of alanine could indicate a disruption in the citric acid cycle, glycolysis and/or glucogenesis.

The bucket region 2.35 - 2.37 ppm is attributable to pyruvate, which is elevated in GD plasma profiles. Pyruvate is a product of glycolysis and may be upregulated in view of a lack of/disturbance with pyruvate dehydrogenase in this disease. This could also serve as a potential marker in GD1, in combination with other metabolic markers described herein.

The bucket region 2.20 – 2.25 ppm is attributable to acetone and a $C\underline{H}_2CO$ resonance which appear elevated in the GD cohort and was statistically significant in one of the multivariate tests. Acetone may represent a significant metabolite, as generally lipids are downregulated in GD and high levels of acetone are often correlated with diabetes and ketoacidosis.

In rat serum it was demonstrated that anaemia has an impact on metabolic pathways [149]. Indeed, an elevation in vLDL cholesterol, LDL cholesterol, acetone, tyrosine, valine, histidine, lipid and TMAO levels and a reduction in succinate, choline, creatine, glutamine, arginine and α-ketoglutarate was observed in rat serum [149]. Some of these metabolite irregularities correlate with our findings since increased tyrosine and lipid levels were observed, and decreased choline levels too. However, not all metabolites are changing in the same manner as those observed in [149].

Metabolites linked to Parkinson's Disease, Niemann Pick Disease Type C and Gaucher's Disease

It has been previously proposed [2], [143] in the literature that an improved understanding of GD1 would enable a better understanding of Parkinson's Disease (PD) since frequently, carriers of GD alleles and GD patients are at a higher risk of developing PD. Hence, metabolites that were significant biomarkers in GD1 with those observed for PD were compared, in order to distinguish any differences or similarities.

Pyruvate and myo-inositol found to be elevated and citrate and acetate were depleted in both GD1 and PD [150]. Other metabolites which were significantly different in PD included elevations in glucitol and propylene glycol, and depletions in galactitol, ascorbate, threonate, glycerol, methylamine, ethanolamine, suberate, isocitrate, glutarate, trimethylamine and malate [150]. Upregulations in pyruvate have also been observed in the CSF of Alzheimer's patients [151], [152]. Pyruvate has been suggested as a marker in PD1, and it was demonstrated that 46 genes interact with the pyruvate dehydrogenase, but two in particular were established as those linked to this disease, namely NPFF and PDHB [150]. The former gene, NPFF is associated with inflammation modulation, neuroendocrine function and cardiovascular regulation, whereas the latter is involved in mitochondrial dysfunction and the activity of pyruvate dehydrogenase [150].

Citrate is a metabolite that is part of the TCA cycle, and is also readily available in the diet [99]. Citrate (2.63 – 2.69 ppm) was depleted in GD1, which could be attributable to several reasons. Citrate is observed to rise upon fasting in plasma profiles [153]. However, another hypothesis is that the depletion of citrate correlates with the noted disruption of pyruvate dehydrogenase [150]. A final hypothesis links citrate to bone metabolism, whereby hypocitricemic conditions could be linked to the lack of calcium content in the bones [154], a process giving rise to osteopenia, osteoporosis joint pain, arthritis and joint damage in GD patients [155].

Myo-inositol is produced by the human body from glucose. Therefore, it was removed from the statistical analysis since prior analysis involved comparisons between fasting and non-fasting participants. Nevertheless, the elevation of myo-inositol observed in the initial analysis (data not shown) may be significant in GD pathogenesis. In PD, it was observed

that myo-inositol was elevated in the basal ganglia under exercise conditions [156]. This upregulation could be relevant, since it has been previously hypothesised that the polyol metabolic pathway could thus be causing mitochondrial dysfunction [150], which has been previously demonstrated to occur in GD1.

Interestingly, metabolic perturbations that have been demonstrated in Niemann Pick Disease Type C1 (NPC) have also been observed in GD in this work [26]. In NPC1 disease, it was observed that plasma HDL is downregulated [26], which is also observed in GD, suggesting that potentially this is an unspecific biomarker for both LSDs. It is evident and clear that lipid disturbances are a feature in both disorders.

3.6.2 Potential Biomarkers for GD1 in Urinary Profiles

From the results acquired, it can be noted that there are no potential biomarkers for urine to be further considered. It is promising that the impact of GD1 enzyme replacement therapy may be providing ^{1}H NMR profiles that are similar to the controls, and this may provide a reason as to why GD biomarkers in the urine are not observable, since all patients were undergoing some form of treatment. However, without having a substantial cohort of untreated GD participants it would not be possible to confirm if the urinary metabolic status resembles that of the control or treated GD cohort. Additionally, in view of the comparison of fasting and non-fasting samples, there may be a more significant impact if both cohorts were collected under the same, preferably non-fasting, conditions.

Interestingly, no bile acids (BAs) between 0.66 – 0.69 ppm (Figure 3.6) were observed as reported in NPC1 [13]. This was described previously as being possibly ascribable to liver trauma [13], and NPC1 and GD1 conditions exhibit similar symptoms, so it was hypothesised that a similar biomarker patterns in the urine may be found. No BA pattern that was distinct to GD1 was observed, and it was hypothesised upregulated BA concentration may have arisen from inflammation of the liver. This is promising for the NPC1 investigation as this may indicate that this specific bile acid pattern is unique to NPC1 and cholesterol disturbances. A previously unassigned δ = 0.76 ppm *iso*-propyl-CH_3 doublet, was observable in both controls and GD1 participants. However, BAs are often present in low concentrations in healthy human urine anyway and are directly impacted as a result of fasting or non-fasting regimens.

3.6.3 Xenobiotics

A combination of xenobiotics assigned, which were found in human urine in this study are listed in Appendix 4. Additionally, with patients taking 30 different medications, this renders metabolomics studies extremely challenging. On sample collection, most GD1 patients were on an average of 4 different types of drugs, with the range of medications including enzyme treatment, antibiotics, analgesics etc. Thus, this makes it challenging to effectively monitor metabolic responses arising from the ERT alone, for example. Henceforth, the complications of drug administration to the metabolomics analysis of human urine is outlined. Moreover, conditions (co-morbidities) that are related to the drugs will be explored, since such conditions may exert an impact on the results. All controls patients were not receiving treatments, unless otherwise specified.

Some of the participants in the study were receiving drugs for GD treatment. This includes the enzymatic drugs imiglucerase (n = 14) and velaglucerase alfa (n = 11) which tackle the disease by means of ERT; with elevated enzyme, the glucosylceramide substrate is reduced in concentration. Eliglustat (n = 1) partially inhibits the enzyme glucosylceramide synthase producing less glucosylceramide. Some participants (n = 5) in this study were not receiving enzyme-related treatment. As shown in the previous Figures, metabolomics investigations were unable to distinguish between participants on different types of enzyme-related medication. However, it remains a possibility that all the treatments were working efficiently, and therefore no metabolic changes in the urine were observed. Some changes in urinary profiles may have been expected in view of previous literature reports which have shown that urinary lipid metabolites, namely monohexosylceramides, showed significant differences when compared to those of controls [157], an observation suggesting disturbances in the lipid profiles of these participants. However, these elevations in monohexosylceramides were found to decrease upon treatment, and their levels returned to normal [157]. The lipid concentration in the urinary profiles was not observable by ^{1}H NMR analysis in this study since it is below the limit of detection. With only 5 participants untreated with some form of GD1 medication, it is very difficult to create a urinary metabolomics medication sub-set in view of a small sample size in each group. The significance drawn from this sample size might not be accurate or precise, and could be subjected to multivariate models that overfit datasets. Additionally, the true effects of eliglustat are also difficult to evaluate with only one participant in this group.

A number of participants were also receiving antibiotics including clarithromycin (n = 2), phenoxymethylpenicillin (n = 7) and azithromycin (n = 1). From our studies in feline urine (Chapter 4) azithromycin was undetectable in any of the urinary profiles examined, since this particular antibiotic is excreted in the bile and kidneys. Similarly, phenoxymethylpenicillin undergoes metabolism through the kidneys. Other penicillins have been observed in human urinary profiles by NMR analysis [158]. Clarithromycin undergoes hepatic metabolism [159], and from the work performed on feline urine, it was not possible to identify this drug in the urinary profiles. Moreover, no antibiotic resonances were identified in urine profiles in the GD1 patient spectra. Participants receiving these drugs did not appear to be 'outliers' in either of the plasma or urinary multivariate scores plots.

Salbutamol (n = 1) and beclomethasone (n = 1) were also used as treatments for GD patients, and are both used for managing asthma. NMR metabolomics profiling has been previously performed on urine samples collected from patients with no asthma, stable asthma and unstable asthma [160]. Urine metabolites which were critical in the model were those related to the TCA cycle inclusive of 2-oxaloglutrate, succinate, fumarate, 3-hydroxy-3-methylglutarate, threonine, *cis*-aconitate and *trans*-aconitate [160]. This cohort did not control the time of urine collection, nor diet before collection, which may have impacted the overall results acquired. Nevertheless, samples from GD1 patients receiving these medications were not 'outliers' in the present study comparing GD1 and controls. It is unknown how drugs, such as salbutamol and beclomethasone, impact on the urinary and plasma profile in GD1 participants and asthma patients more generally, since such metabolomics studies are yet to be performed.

Analgesics are often prescribed to GD1 patients since it is a multiorgan disorder which varies in severity. The GD1 participants were receiving the following therapies: etoricoxib (n = 1) and fentanyl (n = 1); co-codeamol, paracetamol and codeine (n = 2); co-dydramol, dihydrocodeine tartrate and paracetamol (n = 1); acetaminophen (n = 5); aspirin (n = 1); and morphine (n = 1). Control participants (n = 2) were also receiving acetaminophen treatment.

Both aspirin and acetaminophen have been previously analysed in urine using ^{1}H NMR analysis [113]. The impact of the resonances from these drugs and their metabolites appearing in the ^{1}H NMR spectra (Figures 3.14 and 3.15) unfortunately precludes these spectral regions in view of their overlap or potential overlap with those of endogenous metabolites. Aspirin (O-acetylsalicylate), which can be used as an analgesic, can also be used

for cardiovascular disease or stroke. Neither O-Acetylsalicylate, nor its metabolites, were not observed in the ^{1}H NMR plasma profiles.

Tramadol (n = 1) was employed to treat a single GD patient. Its metabolism have been previously extensively explored, and it has several metabolites, including O-desmethyltramadol, N-desmethyltramadol and N,O-desmethyltramadol, which are formed in the liver [161]. Tramadol has been previously observed in the urine of dogs and cats using mass spectrometry [162]. Despite extensive analysis of tramadol's metabolic pathway, it is unclear how it impacts upon the metabolism of humans, if at all.

A single GD patient received morphine treatment. Metabolomics studies focused on morphine administration have been performed in rats, and significant changes were observed in the urine and plasma [163]. Rat plasma was observed to change after morphine addiction, with depleted levels of 3-hydroxybutyrate, L-tryptophan and cystine, and elevated levels of *n*-propylamine [163]. The GD patient receiving morphine therapy was not found to be statistically significantly different in the plasma profiles comparably to the remainder of the GD cohort. Many metabolites in the rat urine changed as a result of morphine administration and morphine addiction, including TCA and amino acid metabolites, amongst others [163]. However, no significant differences were observed in this GD patient's urine *versus* control urine samples. Overall, analgesics are an essential part of managing pain and palliative care in GD. This creates challenging human metabolomics models for GD1.

Omeprazole (n = 2) and esomeprazole (n = 1) were taken by GD1 participants, and are frequently used to treat gastroesophageal reflux disease (GERD), and to avoid side-effects arising from strong pain relief.

In a previous study, human patients with cirrhosis showed different metabolic urinary patterns to corresponding controls, and the differences observed included hippurate, dimethylamine and lactate [164]. On administration of omeprazole, hippurate was significantly reduced in urinary profiles of cirrhosis patients [164]. Additionally, lactate was observed as elevated in the urine of both controls and patients when compared to participants receiving omeprazole [164]. Finally, dimethylamine was seen to be significantly reduced in the urine of the cirrhosis cohort [164]. However, in this study, no significant differences in the GD1 profiles arising from the administration of omeprazole was observed in multivariate testings in plasma but an outlier was observed in urine. The study reported in (63) did not analyse human plasma samples, and to date the metabolic differences in human plasma have not been observed using omeprazole. However, one study which analysed blood

serum samples in rats following administration of aspirin and a combination of aspirin and omeprazole [165] observed no disturbances in the citric acid cycle, but did find some disturbances in beta-oxidation with a depletion in 3-D-hydroxybutylate [165]. No significant differences in the plasma of the GD1 cohort were observed, and these samples were not observed as outliers.

Methotrexate (n = 1) has been used to treat a GD1 patient in this investigation. Methotrexate has previously been explored in a ^{1}H NMR metabolomics investigation which evaluated the treatment outcome and the corresponding metabolic changes in urinary profiles in 30 rats [166]. A total of 13 biomarkers for successful treatment were proposed, including urate, taurine, histidine, methionine, glycine, tryptophan, uracil, TMAO, methionine, aspartate and α-oxoglutarate [166]. Despite this, the GD1 participant being treated with methotrexate did not exhibit any 'outlier' characteristics in multivariate scores plots of the plasma or urinary datasets.

Benzydamine (n = 1), ibuprofen (n = 1) and naproxen (n = 1) have been used to treat GD1 patients. Ibuprofen and its corresponding metabolites have been well documented in human urine [115], [167]. However, ibuprofen metabolites were not observed in the urinary profiles of GD1 patients, and this may be because of the drug's rapid metabolism (half-life *ca.* 2 hr). In the metabolic profiling of rat urine using ^{1}H NMR analysis, naproxen was found to induce metabolic increases in creatine, creatine phosphate, betaine and kynurenate, and a reduction in choline [168]. Additionally, naproxen dosage depends on how significant these differences are. For example, significant changes in citrate levels are only observed on comparison of vehicle *vs.* high dose-administered rats. In this study, no significant differences between GD urine in multivariate testing was observed, and nor were there any 'outliers' arising from the taking of non-steroidal anti-inflammatory drugs.

Some patients with GD were also taking bendroflumethiazide (n = 1) candersartan (n = 1) hydrochlorothiazide (n = 1) losartan (n = 1) and co-amilofruse (n = 1), which is a combination of amiloride [169] and furosemide. Losartan metabolism has been investigated using ^{1}H NMR in human plasma profiles, and showed a metabolic response in biomolecules such as upregulated vLDL/LDL lipids, creatine, creatinine, lactate, citrate and α-glucose in addition to downregulated HDL lipids, glycine, phosphorylcholine and choline [170]. These markers were indicative of pharmacometabolomic effects arising from this treatment. Moreover, losartan was transformed to its bioactive metabolite EXP3174 [170].

Citalopram (n = 1) and amitriptyline (n = 3) were taken by small numbers of the GD1 patients in the cohort. Selective serotonin reuptake inhibitors (SSRI's) have been previously subjected to metabolomics analysis and elevated glycine has been proposed as a response biomarker for these conditions in human plasma samples [171]. However, the study did not explore the biomarkers as a response to treatment outcome and if glycine was depleted as a result of such administration, which is required for validation purposes. A ^{1}H NMR metabolomics study has been performed on the urinary profiles of patients receiving citalopram, and it was found that the metabolic disturbances caused by major depressive disorder returned to control levels [172]. Twenty-four out of twenty-six urinary metabolites monitored by UPLC-Orbitrap-MS have also been observed to return to normal levels after administration of citalopram in rats with mild stress [173]. It would be expected metabolic disturbances would be observed in urine as a result of administration of citalopram herein. No significant differences in the GD1 urinary profiles were observed; however, this is a consideration that should be observed in future investigations. Glycine was not observed to be a significant metabolite in the plasma profiles in these particular GD patients, nor did this GD patient appear as an 'outlier' in multivariate analysis scores plots.

Quinine sulphate (n = 1) has been administered in a GD1 participant in this study, and is often used for muscle cramps. Quinine sulphate is metabolised in the liver and excreted in urine, and a total of 10 metabolites have been previously identified by GC-MS in human participants [174]. Quinine sulphate can cause hypoglycaemia, but regardless, glucose was removed from the multivariate analysis performed on plasma samples, and this participant's sample was not an 'outlier' in this metabolomics investigation.

Pregabalin (n = 1) is a treatment that manages a range of conditions including epilepsy, fibromyalgia, restless leg syndrome and generalised anxiety disorder, and was administered to one of the GD1 patients. This participant, however, did not appear as an outlier in multivariate analysis scores plots for the plasma profiles.

Warfarin (n = 1) was used to treat one GD1 participant. Plasma has been analysed in participants receiving warfarin and it demonstrated that ethanolamides, a group of lipids responsible for procoagluation, were significantly upregulated [175]. This was not observed in this investigation, and this sample did not appear as an 'outlier' amongst the plasma samples.

Zoledronate (n = 2) is used to treat various bone diseases, including osteoporosis, which is frequently prevalent in GD1 patients [176], amongst other bone-related conditions. No

resonances attributable to this therapeutic agent in the plasma profiles acquired. Zoledronate is metabolised by the kidneys and excreted in urine without any structural changes, although it was not observed in the ^{1}H NMR spectra.

Loperamide is an opioid-receptor agonist used to treat diarrhoea and was prescribed to treat a GD participant in this cohort. Loperamide is only excreted in small amounts in the urine and is considerably metabolised *via* first-pass metabolism. Serum biomarkers have been found in rats that are indicative of loperamide ingestion [177]. Depletion of acetate, glucose, glycerol, lactate, succinate, taurine, alanine, glutamate, glutamine and glycine were observed in the blood serum of loperamide-treated rats [177]. Hence, administration of this particular drug may impact the results of the plasma acquired in this study for this participant. However, this participant did not appear as an 'outlier' in any of the metabolomics scores plots generated. Additionally, metabolic changes demonstrated in rats were not observed in this human trial.

Some of the GD participants were on cholesterol-lowering drugs, namely simvastatin (n = 2) and pravastatin (n = 1). For participants to be prescribed these drugs, they are likely to have high cholesterol levels, and potentially be overweight with high BMI values, and are older and/or have a higher risk of cardiovascular disease, specifically stroke, heart attacks, coronary heart disease and angina. All of the above conditions and factors are required to be taken into account prior to a metabolomics investigation; however, this information was not provided for the GD1 group.

In a metabolomics study monitoring rat plasma by ^{1}H NMR, simvastatin-treated rats were observed to have higher levels of LDL/vLDL-TAGs, 3-D-hydroxybutyrate, glutamine, O-acetyl glycoproteins, lactate, choline/phosphocholine, and depleted levels of alanine, polyunsaturated fatty acids, glutamate, arginine, monounsaturated fatty acids, valine, proline and pyruvate, when compared to corresponding controls [178]. Other studies have been performed in order to assess the impact of weak and high simvastatin toxicity with ^{1}H NMR monitoring of rat urinary profiles [179]. The biomarkers found include allantoin, 2-oxoglutarate and TMAO, which were all increased in the urine upon treatment with simvastatin [179]. Since the present GD1 urinary metabolomics investigation was insignificant, this may contribute towards further overlap of multivariate scores between controls and GD1 participants.

Serum cholesterol levels were also demonstrated to be lowered in pravastatin-treated participants [180]. Additionally, changes were observed in apolipoproteins B and a

combination of these with apoliproteins A1. Statins therefore may have a positive impact on GD1 participants as the disorder is a LSD, and lipids for the most part are found to be lowered on administration of statins [181]. This could exacerbate the significance of what is observed in GD1, since if plasma cholesterol is low in GD1 participants, and it is further lowered upon receiving pravastatin, this might explain some of the metabolomics analysis overlap between control participants and GD participants. However, in view of the small sample size of the number of participants taking statins, a full metabolomics investigation would have to be performed in order to validate this hypothesis.

Participants in the GD1 group were taking estradiol (n = 2) and norethisterone (n = 1), whereas healthy controls were not receiving any hormone treatment or contraceptives. ^{1}H NMR metabolomics investigations have been previously performed in plasma and urinary profiles, observing the effects of menstrual cycle phase, which found no separation based on urinary metabolites [182]. There were a range of participants who participated in this study, and these included a variety of ages, and some of these participants were taking oral contraceptives [182]. However, these participants did some show variations in plasma between two phases of the menstrual cycle, with the menstrual and luteal phases being significantly different that were significant [182]. This included a significant decrease in metabolites such as glutamine, glycine, alanine, lysine, serine and creatinine, and elevations in acetoacetate and vLDL-TAGs [182]. On comparison with data collected here, only alanine is depleted in GD1 plasma, and no significant changes in any other metabolites were observed. Elevation in the creatinine/creatine resonance was also observed in the GD1 patients. However, this observation differs from the depletion of metabolites observed in the menstrual cycle study. [182]. In a study that compared sex and contraceptive use, it was shown that women receiving oral contraceptives showed lower levels of glycine, proline, histidine, lysine, hydroxyproline and orthenine, along with increased levels of isoleucine [183]. GD datasets did not exhibit any of the changes described above in the plasma profiles. No outliers were observed arising from hormone treatment.

Biofluid	Drug and dose	Metabolites	Pathways disturbed and impacted metabolism	GD1 metabolites
Plasma[b]	Morphine addiction	3-Hydroxybutyrate (↓), l-tryptophan (↓), cystine (↓), n-propylamine (↑)	(↓) β-Oxidation pathway from fatty acids and/or ketone production from acetyl-CoA Some biological implications unknown	n/a
Urine[b]	Morphine addiction	(↑) 2-Ketoglutaric acid, fumaric acid, malic acid and l-threonine (↓) Glutamate, isoleucine, valine, aspartic acid, oxamic acid, 2-aminoethanol, indoxylsulphate, creatinine		n/a
Urine[b]	Naproxen Toxicity	(↑) Creatine, creatinine phosphate, betaine, citrate, cis-aconitate, kynurenate (↓) Choline	Energy metabolism decreased mitochondrial function and gastric injury	n/a
Plasma[a]	Losartan	Glycine, phosphorylcholine, choline, creatine, creatinine, lactate, citrate alpha-glucose and lipids	Glycine, serine and threonine and glycolysis or gluconeogenesis metabolism	
Plasma[b]	Simvastatin	(↑) LDL/vLDL-TAGs, β-hydroxybutyrate, glutamine, O-acetyl glycoproteins, lactate, choline/phosphocholine (↓) Alanine, polyunsaturated fatty acids, glutamate, arginine, monounsaturated fatty acids, valine, proline pyruvate	Lipid metabolism ketone bodies gluconeogenesis	(↓) LDL/vLDL
Urine[b]	Simvastatin toxicity	(↑) Allantoin, 2-oxoglutarate, TMAO	Free radical damage, oxidative stress, inflammatory response	n/a
Serum[a]	Pravastatin	(↓) All cholesterol types except HDL*	Reversal of cholesterol transport	(↓) All cholesterol
Serum[a]	Warfarin	(↑) Ethanolamides	Procoagluant	n/a
Serum[b]	Loperamide	(↓) Alanine, glutamate, glutamine, glycine, acetate, glucose, glycerol, lactate, succinate and taurine	TCA cycle	n/a

Table 3.4. Impact of administration of drugs on biofluids, inclusive of what metabolites are affected in humans[a] and rats[b] and considering a combination of NMR- and MS-determined metabolites. Corresponding metabolic pathway that are disturbed are also shown. *scaled against a baseline measurement rather than corresponding control participants.

Drug metabolite(s)	Chemical shift (δ/ppm)	Urine metabolite(s)
Acetaminophen[a]	2.15 6.90 7.25	Glutamine-CH_2 3-(3-Hydroxyphenyl)-3-hydroxypropanoic acid-CH
Acetaminophen glucuronide derivative[a]	2.15 3.63 3.91 5.11 7.13 7.34	Glutamine-CH_2 Myo-insitol-CH/Glycerol-CH_2 Creatine-CH_2 Tartrate-CH
Acetaminophen sulphate derivative[a]	2.18 7.27 7.45	Acetone-CH_3 Indoxyl sulphate-CH Imidazole-CH
Acetaminophen L-cysteinyl derivative[a]	1.84 2.15 3.28 4.30 7.50	Aminoadipate-CH_2 Glutamine-CH_2 Taurine-CH_2 Choline-CH_2 Indoxyl sulphate-CH
Acetaminophen N-Acetyl L-cysteinyl derivative[a]	2.15 3.35 3.99 6.99 7.26 7.51	Glutamine-CH_2 Methanol-CH_3 Phenylalanine-CH 3-Methylhistidine-CH Indoxyl sulphate-CH Indoxyl sulphate-CH
Salicylurate[a]	3.96 3.97 7.02 7.49 7.78	 Phenylalanine-CH Histidine-CH Hippurate-CH

Table 3.5. A summary of drug resonances found in human[a] and animal[b] urinary profiles in this study and the corresponding bucket region elevated and the corresponding urinary metabolites which would not be able to be quantified in view of interfering drug/drug metabolite resonances.

3.6.4 Diet and Dietary Supplements

Diet has an impact on ^{1}H NMR metabolomics and this has been shown in various studies [184]. Participants in the control group included a range of diets which were documented using a food diary. None of the participants in the control group had taken supplements with

the exception of one participant taking magnesium supplementation. GD participants were taking vitamins and this can impact the metabolic pathways. Herein, it is described what supplements were taken and how this can potentially have affected the results.

Cod liver oil is a very well-known supplement and has been previously evidenced for being beneficial for heart disease in population studies [185]. However, metabolomics studies have not been performed on cod liver oil and there is some evidence that suggests cod liver oil has no influence on disease [186]. Nevertheless, it has been demonstrated that seafood diets give rise to a higher level of TMAO in urinary profiles whereas, guanidinoacetate and 3-methylhistidine are reduced in non-seafood diets [187]. This may have an impact when conducting a metabolomics investigation whereby some of the GD participants (n = 2) are taking cod liver oil supplements, known for improving cardiovascular health. Cod liver oil is known to lower total cholesterol and LDL cholesterol levels in the blood, which is believed to be due to elevated expression of cholesterol receptors [188]. A depletion in cholesterol in plasma was observed across the whole GD cohort and not just the two GD participants receiving cod liver oil.

Vitamins are frequently taken by many groups of diseased cohorts as they can be taken due to a deficiency and other medical reasons such as pregnancy or digestive disorders. It can be difficult to isolate the impact of a vitamin on the metabolism and often vitamins are used in combinations with other minerals, such as calcium, magnesium, and drugs.

Folic acid, folate or vitamin B9 (n = 2) is another supplement taken by the GD cohort. The intake of folic acid has been shown to significantly elevate plasma homocysteine and urinary malondialdehyde and lymphocyte DNA hypomethylation in postmenopausal women [189]. Elevation in plasma homocysteine can lead to an elevation in taurine as demonstrated in rats [190]. Additionally in diabetic rats, 20 urinary metabolites were found to be significantly altered as a result of folic acid supplementation including urinary metabolites such as creatine, nicotinurate, n-acetyl-tyrosine spermine, 5-hydroxyindoleacetate, Indolylacryloylglycine, 4-(2-Aminophenyl)-2,4-dioxobutanoic acid, hydroxyphenylacetylglycine, suberic acid, 2-Oxo-4-hexenoic acid, 5-oxo-7-decenoic acid, ω-oxo-capric acid, ethyl oxalacetic acid, hydroxy-cyclohexenoic acid, monoethylhexyl phthalic acid and cis-stilbene oxide [191]. These metabolites demonstrated potential changes in the gut microbiota [191]. No significant differences in the patients GD urine were observed, across 30 different patients, these metabolic changes observed in the literature are unlikely to make a significant change to the study. Although a potential elevation in taurine in plasma was observed, these buckets were

shared with those of glucose and required to be removed from the investigation, thus do not impact the plasma results herein. Other B vitamins that were taken in this study by the GD cohort include cyanocobalamin (n = 1), thiamine (n = 1) and hydroxocobalamin (n = 4). Overall, these B vitamins boost energy metabolism but the participants do not appear as outliers in the statistical testing.

A number of vitamin D's were taken in the GD group including ergocalciferol, vitamin D_2, calcitriol, vitamin D, calcium carbonate and vitamin D_3. Vitamin D is often taken in order to strengthen teeth and bones and therefore in GD, where the symptoms are bone pain for example, it is an important vitamin to be supplementing. To date, no metabolomic studies have been performed on vitamin D that could explain some of the changes observed herein.

One participant in the control group was taking magnesium vitamin supplement. This participant was not observed to be an outlier in the statistical tests on urine or plasma profiles suggesting no dramatic changes in metabolism. Although this could be further investigated in a larger metabolomics investigation as magnesium is observed to play a role in energy metabolism.

3.6.5 Implications of comparing fasted with non-fasted participants plasma

Many studies have previously explored the effect of fasting and non-fasting participants and it has been shown that even acute dietary changes can affect the metabolism [192]–[194]. Glucose and amino acids have been shown to give a difference between fasting and non-fasting participants [194]–[196]. Thus, it is crucial further experiments are performed with fasted GD participants to validate these findings to ensure a rigorous metabolomic study.

Removal of buckets in plasma profiles pertaining to glucose, phenylalanine and myo-inositol was important, as although these metabolites could be due to clinical significance, they are more likely to be upregulated due to simple differences between fasting and non-fasting participants [197].

Glucose was elevated in the GD participants plasma (Table 3.3), which is expected, due to the lack of fasting from these participants in comparison to the healthy controls [102], [122]. Glucose is a precursor for myo-inositol, if glucose is elevated, consequently further myo-inositol is produced in a two-step reaction *via* glucose-6-phosphate [148]. Phenylalanine is also expected to be elevated due to being an essential amino acid that is required to be

obtained from diet and therefore is likely to be of little clinical significance [102]. Phenylalanine converts to tyrosine by phenylalanine hydroxylase, hence why both of these metabolites may also be increased in the plasma. Tyrosine however, can also be synthesised by the body similar to formate, both of which are elevated. Tyrosine is not found in high concentrations throughout the body, as normally it is rapidly metabolised and one study suggests it is higher in fasting individuals as opposed to non-fasting individuals [102]. However, tyrosine is elevated in GD patients, the non-fasting individuals in this case. Large concentrations of tyrosine usually indicate a need for more vitamin C, which is crucial to bones and in GD, bones are impacted as a result of the disease.

Histidine is another essential amino acid which is elevated in GD. This metabolite in serum from fasting participants has been shown to be both higher [102] and lower [122] in fasting participants however, the former is evident here. Histidine is attributable predominantly from dietary sources but elevated levels of plasma histidine have been previously associated with numerous pathological conditions such as tremors, ataxia and psychosis [148]. Therefore, it is challenging to identify herein whether this metabolite is elevated due to the different experiment conditions or alternatively the disease pathogenesis.

Leucine is a precursor of 3-HB and is elevated typically in fasting participants serum [102] and it is an essential amino acid which can only obtained through the diet. Thus, an elevation in leucine would cause the elevation of 3-HB. Leucine is elevated in our statistical analysis performed (Table 3.3), which could be due to comparing non-fasting and fasting participants. Leucine has not previously been described or identified as a marker for GD.

Fasting is known to have an effect on urinary metabolites including an increase in 3-HB [60]. Furthermore, diet is also shown to have an effect on metabolites within urine; mannitol is said to increase with the intake of carrots, asparagus and apples [60], trans-ferulic acid increases with the intake of breakfast cereals and chocolate [60], betaine/proline decreases with low prevalence of dietary fruits [60], [184], dimethyl sulfone increases with intake of onions [60], methyl histidine increases with intake of meat [60], ethyl glucoside is elevated following the ingestion of rice, wine and sake [184]. Hence, it is important to take into consideration these other metabolites which may be impacted.

Urinary metabolic profiles can also be impacted by drinks as well as food. Urinary profiles have been correlated with age and this has previously been demonstrated due to a higher intake of caffeine [112], [119]. No significant differences in the profiles were observed herein and elevated trigonelline levels were not reported in the GD participants.

3.6.7 Populations

The GD cohort and cohort of participants from DMU were different populations. The INTERMAP study highlights the impact of population differences in metabolome inclusive of changes in gut microbial-host metabolites, amino acids, dietary related metabolites, energy metabolism, and TCA intermediates in urine [184]. This can be correlated with systolic and diastolic blood pressure [184], which was not taken as part of the study herein but it is important to consider in future metabolomics investigations. Not only were differences found between populations such as East and West of the world but also within populations [184]. The north and south of China had significantly different urinary metabolites. These were believed to be derived from different dietary patterns, blood pressure and BMI and this is reflected in the occurrence of CVD and mortality rates [184]. This study explored the impact of environmental differences which could be potentially why differences are observed between GD1 and control cohorts.

3.7 Conclusions

A clear distinction between GD and control plasma profiles has been established using ^{1}H NMR combined with MVA. There are evident lipid disturbances in plasma profiles which have been previously reported in the literature that corroborate with the findings in this study. Furthermore, upregulated metabolites such as lactate, creatinine and depletion in tyrosine and phenylalanine may indicate liver dysfunction.

This ^{1}H NMR metabolomics strategy is able to easily distinguish from control and GD1 patients, despite the fasting period. However, it is important to consider the drawbacks and implications of these findings due to comparison of fasting and non-fasting participants in addition to GD1 patients being on a plethora of drugs which are highlighted herein.

No distinction was made between urinary profiles in controls and GD1 patients using ^{1}H NMR. These participants were receiving enzyme treatment, which has previously been shown to affect the urinary lipid metabolome in GD, but no other pathways herein have been explored using a ^{1}H NMR approach, such as the amino acid metabolism for example. Despite this the results are interesting as, the potential effectiveness of enzyme treatment in

GD is demonstrated, as the multivariate and univariate analysis is not statistically significant in GD profiles compared to the control profiles.

Herein, it was also explored how drugs also present challenges in human metabolomic studies as some elements of experimental design cannot be controlled such as drug administration. Moreover, metabolomics studies have not been performed on all drugs and predominantly animal models have been used to investigate the impact of drugs on metabolites. Animal models allow for uniformity in experimental design which is not always possible in human studies which makes exploring drugs using these models practical. It is also challenging to isolate the impact of a single vitamin or supplement when extensive studies have not been performed on corresponding controls or the disease group. It is very well known that diet and nutrition can impact metabolomics outcomes but, due to a limited sample sizes, these issues were not explored further.

3.8 Proposal for future research investigations

From this work complications of uncontrollable human variables into a metabolomics investigation are described. This creates complexity in the models and variance can be due to a number of factors including treatment, age, sex etc. It would be essential to replicate this study with an animal model with treated, untreated and control animals whereby it is easier to control the conditions exposed to the subjects. Alternatively, it would be desirable to have samples from fasted GD1 participants where possible on no medication that would cause interference with the spectrum. However, due to the patients being in critical condition in some instances this may not be possible, but this introduces further variables in a metabolomics investigation including the aforementioned drug signals present in spectra and the changes in metabolism. Another investigation that would validate findings herein would be to observe human plasma and urine samples from GD1 participants prior- and post-ERT or other therapies such as Eliglustat.

Chapter 4: A ^{1}H NMR Metabolomics Investigation of Niemann Pick Disease and Hydroxypropyl-β-Cyclodextrin Therapy in an Animal Model System

4.0 Abstract

Background: Niemann Pick Disease Type C (NPC) is a rare, neurodegenerative, lysosomal storage disorder which alters the storage and transport of cholesterol and sphingolipids. Hydroxypropyl-β-Cyclodextrin (HPβCD) has been proposed as a treatment for NPC in view of its cholesterol encapsulation and removal properties.

Methodology: A ^{1}H nuclear magnetic resonance (NMR) metabolomics investigation was performed to assess metabolic responses from felines suffering with NPC, along with the metabolic effect of HPβCD treatment. Urinary profiles were acquired using a 400 MHz NMR spectrometer analysing control (n = 35), untreated NPC (n = 22) and HPβCD- treated NPC felines (n = 95). Three HPβCD treatment approaches were used: two intrathecal methods were administered using a 120 mg dose, one initiated at 3 weeks pre-symptomatically (n = 36), and another initiated at 6 weeks, post-symptomatically (n = 36). The third approach involved subcutaneous injection of 1,000 mg/kg and 120 mg doses intrathecally, and this was initiated at 3 weeks of age (n = 23).

Results and Discussion: More than 80 metabolites were detectable in the feline urine spectra, including short-chain organic acid anions, branched chain amino acids, N-acetylamino acids, N-acetylsugars, and notably nicotinate and nicotinamide pathway metabolites, amongst others. Partial Least Squares-Discriminant Analysis (PLS-DA) was unable to distinguish between control and NPC felines. However, on removal of the younger felines (< 10 week ages), in which disease progression was not as established and hence pre-symptomatic, PLS-DA was able to successfully distinguish between these restricted groupings. Significant differences were not found when comparing the treated and untreated NPC felines using the intrathecal 120 mg dose both pre- and post- symptomatically. However, for felines treated by subcutaneous injection (SQ) of 1,000 mg/kg and intrathecal injection (IT) of 120 mg of HPβCD initiated at 3 weeks of age, significant differences were observed. Significant metabolites included felinine, creatine/creatinine, hippurate, phenylacetylglycine, trimethylamine oxide, betaine and taurine, together with others. These urinary results were compared to results obtained on blood plasma in order to assess the

outcomes of different treatment models, and further understand the metabolic pathways involved.

Conclusions: ^{1}H NMR metabolomics profiling can successfully distinguish between treatment approach outcomes in the ^{1}H NMR urinary profiles of felines. ^{1}H NMR metabolomics is unable to distinguish between NPC feline samples and controls; however, post-symptomatic changes can be observed when evaluated against control felines. Treatment with a combination of SQ (1,000 mg/kg) and IT (120 mg) of HPβCD initiated at 3 weeks of age was deemed to be the most effective course of treatment. Removal of the SQ 1,000 mg/ kg and IT 120 mg HPβCD treatment group from the complete dataset did not reveal any significant urinary metabolic differences between the other treatment regimens and untreated controls. Metabolic modifications observed in the post-symptomatic NPC urinary profiles indicate disturbances in both gut microbiota and cholesterol/sterol metabolism. These disturbances appear to be reversed on treatment with SQ (1,000 mg/kg) and IT (120 mg) injections of HPβCD.

Keywords: Nuclear Magnetic Resonance, Hydroxypropyl-β-Cyclodextrin, Niemann Pick Disease Type C1, Lysosomal Storage Diseases, Urinary Analysis, NMR-Linked Metabolomics, Metabolism.

4.1 Introduction

Niemann Pick Disease Type C (NPC) (*OMIM 257220*) is a rare, neurodegenerative lysosomal storage disorder which alters the storage and transport of cholesterol and sphingolipids in view of a mutation in either the NPC1 (95%) or NPC2 (5%) genes [2]. Humans and animals afflicted with this disorder present symptoms including, but not limited to, vertical ophthalmoplegia, dementia, ataxia, dystonia and hepatosplenomegaly. These developments are potentially fatal at an early age, which can occur as early as 24 weeks in felines, and 0.1years in humans [198]. The median age of death for humans with NPC disease is 13 years [198]. At present, NPC disease is diagnosed using the Filipin test which is required in duplicate to validate findings, and consists of an invasive skin biopsy procedure [199], followed by a genetic test to confirm the presence of this condition. Miglustat is currently the only approved treatment for NPC disorders, which inhibits the enzyme glucosylceramide synthase, and hence prevents sphingolipid production and slows down lysosomal lipid build-up and hence in turn, neurodegeneration, amongst other symptoms [200]. Such lipid

deposition is not limited to sphingolipids, since cholesterol trafficking is also perturbed in NPC disease, a process leading to substantially higher cholesterol levels.

There is an urgent requirement to seek improved diagnostic and treatment monitoring strategies for NPC and other lysosomal storage disorders. Indeed, NMR-based metabolomics has recently been used in the study of NPC in order to search for new biomarkers [13]. Moreover, the urinary excretion of treatments such as miglustat have been monitored using such an NMR approach [201]. Metabolomics is the study of low-molecular-mass biomolecules under different environments *e.g.* healthy control *versus* diseased humans. Unlike genetics, the metabolome is able to provide a 'snapshot' of the status of the organism at a particular point in time. Therefore, metabolomics is also extremely valuable for the discovering of biomarkers for the monitoring of disease progression and treatments effectivity. Metabolomics methodologies can either be targeted or untargeted; targeted approaches observe specific pre-defined metabolites which are usually known to be related to disease pathology or aetiology, whilst untargeted approaches view all possible global metabolites, including unknown ones. NMR analysis is an untargeted technique, which offers many advantages since it is able to detect and quantify a very wide range of metabolites simultaneously, in order to recognise any unusual molecular 'signatures'. It is also a high-throughput and non-destructive bioanalytical strategy.

HPβCD has been proposed as a treatment for NPC1 disease in animals [202]–[204], with the first human clinical trials on this novel therapeutic agent are now in progress [205]. HPβCD is a cyclic oligosaccharide which has a hydrophilic exterior and a hydrophobic interior; these properties allow this molecule to internally encapsulate lipids such as cholesterol. Dependent upon its concentration *in vivo*, HPβCD is also able to remove (transport cholesterol away) from the body, although the mechanism of this process is not completely understood [206]. Indeed, animal studies are now making it possible to further understand the mechanisms of HPβCD as a therapeutic agent in NPC [207] in addition to Alzheimer's disease [208], and this drug has shown much promise in current trials. However, to date the effect of HPβCD on the human or mammalian metabolomes is unknown. Thus, in this pilot investigation, ^{1}H NMR-based metabolomics was employed in order to identify new, potentially valuable biomarkers for NPC disease in felines, particularly those which correlate with disease severity and progression. Furthermore, this investigation aimed to observe the potential use of HPβCD as a therapeutic treatment regimen in these felines using three different dosing methods. In view of the longitudinal status, urine samples collected were standardised, and repeated urinary collections were made throughout increasing time-periods. In this manner,

the capability of these techniques to prognostically stratify these experimental animals was explored. Indeed, these software-based approaches served to provide valuable information regarding biomolecular connectivities, and hence the biological/metabolic pathways involved in the pathogenesis of NPC diseases.

4.2 Methods

4.2.1 Approval for Research

Cats were raised in the animal colony of the School of Veterinary Medicine, University of Pennsylvania, under NIH and USDA guidelines for the care and use of animals in research. The experimental protocol was approved by the University of Pennsylvania Institutional Animal Care and Use Committee.

4.2.2 Animals and Drug Formulations

Domestic shorthair felines were housed and retained under the same conditions reported elsewhere [204], and were of the same line of breed [204]. The felines were raised at the School of Veterinary Medicine, University of Pennsylvania, under National Institutes of Health (NIH) and United States Department of Agriculture (USDA) guidelines for the care and use of animals in research, within an animal colony. The University of Pennsylvania Institutional Animal Care and Use Committee approved this experimental protocol (NIH OD P40-10939). Guidelines for the care of use in animal research were adhered to, and which are established under the NIH and US Department of Agriculture. All animals were maintained at 21°C with *ad libitum* food and water. 12-hour light cycles were used with air changes 12-15 times per hour. Both classes of HPβCD for IT and SQ administrations were formulated as described in [207], and doses administered are listed below. The dose of 120 mg HPβCD IT and 1,000 mg/kg SQ was determined from the body weight and brain weight of the animals, respectively, as in [204]. Felines were administered with IT injections every 2 weeks, and SQ injections every week, and animals were weighed at 2-week intervals.

4.2.3 Sample Collection

A total of 33 felines were involved in this study, with urine collection occurring at different ages. Animals were fasted for 8 hr. prior to sample collection, and then urine was collected when felines were under propofol anaesthesia for cyclodextrin administration. The feline urine was taken at different increasing intervals, and grouped into 8, 10, 16, 20, 24, 52 and 72 week time-points. For multivariate (MV) analysis, the 50 and 72 week samples were combined in view of a lack of samples at these age ranges, and these were all labelled as 50 weeks. Typically, felines with NPC disease die at a mean age of 21 weeks [207].

Two different types of HPβCD, Sigma for SQ injections, and Kleptose for IT injections, were used throughout these studies. Two sets of felines were treated pre-symptomatically (initiated at 3 weeks of age) using either 120 mg IT HPβCD alone, or a combination of 1,000 mg/kg SQ and 120 mg IT HPβCD injection every 2 weeks. These two treatments are considered clinically the same. A final set of felines were treated post-symptomatically (initiated at 16 weeks of age), using 120 mg IT HPβCD injection every 2 weeks.

Some felines were on complementary medication including antibiotics, which can affect the metabolome. Clavamox (CLX) was used as an antibiotic throughout this study (n = 10 samples), which is a composite of the potassium salt of clavulanic acid (CLAV) and amoxicillin (AMOX). Metronidazole (MET) and Azithromycin (AZI) were also used throughout these studies (n = 4 and n = 3 samples respectively). Full details regarding the metabolism and excretion of these drugs are detailed in Appendix 2.

Sample Type	Number of cats	Male:Female	Samples
Control Untreated	7	4:3	35
NPC Untreated	5	1:4	22
NPC Treated with Cyclodextrin/Kleptose IT 120 mg early (initiated at 3 weeks of age)	6	3:3	36
NPC Treated with Cyclodextrin/Kleptose IT 120 mg late (initiated at 16 weeks of age)	7	5:2	36
NPC Cyclodextrin treated with combination SQ (1,000 mg/kg) and IT 120 mg early	6	4:2	23

Table 4.0. Urine samples collected from feline groups and analysed.

4.2.4 Sample Transportation

Feline urine samples were donated from the School of Veterinary Medicine, Department of Clinical Studies, Philadelphia and abided by the Importation of Animals Pathogen Order 1980 and the Specified Animal Pathogens Order (2008), as per the Council Directive 82/894/EEC (as amended) on the notification of animal diseases within the EU. These were transported to De Montfort University, Leicester on dry ice and immediately stored at -80°C on arrival, until ready for analysis; 152 urine samples were received which are summarised in Table 4.0.

Plasma samples were also collected taken by the University of Philadelphia research group involved, and a clinical chemistry panel was performed on these samples at their School of Veterinary Medicine.

4.2.5 Sample Preparation

Urine samples were thawed at room temperature followed by centrifuging a 500 µl aliquot and removing 450 µl of the corresponding supernatant. Phosphate buffer (Acros Organics) at pH 7.00 (50 µl) was added to the supernatant, in addition to 0.05% (w/v) sodium azide (50 µl) (Sigma Aldrich) and 10% (v/v) D_2O (55 µl) (Sigma Aldrich) containing (w/v) 0.05% 3-(trimethylsilyl) propionic-2,2,3,3-d4 acid (TSP) (Sigma Aldrich) was then added to the solution. This was then thoroughly rotamixed and added to newly-purchased 5-mm diameter NMR tubes ready for analysis.

4.2.6 Sample Analysis

All spectra were acquired using a 400 MHz Bruker Avance I (AVI) NMR (Leicester School of Pharmacy, De Montfort University, Leicester, UK) spectrometer operating at a frequency of 399.93 MHz. The samples were analysed using the water suppression technique (*noesygppr1d*) in order to saturate the water signal (δ = 4.80 ppm) in the urinary samples, with irradiation at the water frequency during the recycle and mixing time delays. 32K data points were acquired using 128 scans and 2 dummy scans, 3 µs pulses, over a sweep width of 4,844 Hz and a receiver gain of 128. The auto-sampler ensured that the samples were continuously delivered throughout the analysis period. Chemical shifts were referenced to TSP, and NMR

spectra were assigned considering chemical shifts, coupling patterns, integral ratios and coupling constants of resonances therein.

Some additional spectra were acquired using a 600 MHz Jeol-ECZ NMR (Leicester School of Pharmacy, De Montfort University, Leicester, UK) spectrometer operating at a frequency of 600.17 MHz. The samples were analysed using the water suppression technique (*single_pulse_wet*) in order to saturate the water signal (δ = 4.80 ppm) in the urine samples, with irradiation at the water frequency during the recycle and mixing time delays. 16K data points were acquired using 64 scans and 4 dummy scans, 6 μs pulses, over a sweep width of 15 ppm and a receiver gain of 36. The auto-sampler ensured that the samples were continuously delivered throughout the analysis period.

Experiments were performed in order to provide confirmatory identification of molecules within the biofluid samples which include the application of 2D-TOCSY and 2D-COSY techniques. These were acquired using the 600 MHz Jeol facility using a presat COSY and TOCSY pulse sequences. The COSY spectra were acquired using 4 pre-scans; 1280 x points and 256 y points; sweep width of 11.28 kHz; an offset of 4.69 ppm; and total number of scans of 2,048. The receiver gain was optimised to 66 and the relaxation delay was 1.5 seconds. The 2D TOCSY was acquired using 4 pre-scans; 1024 x points and 256 y points; sweep width of 11.28 kHz; an offset of 4.69 ppm and a total number of scans 2048. The receiver gain was optimised to 62 and the relaxation delay was 1.5 seconds. Temperature remained constant throughout the duration of the experiments and the samples were not spinning.

4.3 Metabolite Assignments

Metabolites were assigned using the literature available [209], [210], the human metabolome database (HMDB) [65] and Chenomx using the above features of spectra described above. However, at present there is no database for feline metabolites for metabolomics investigations.

4.4 NMR Data Preprocessing for Statistical Analysis

All samples were referenced, aligned, baseline- and phase-corrected. Once this was complete, the samples were then manually bucketed using *ACD Labs 12.0*. The buckets were imported into excel to create spreadsheets containing ^{1}H NMR data matrices. Any buckets attributed to baseline noise, drugs or water were removed. The total number of buckets for the study was 133.

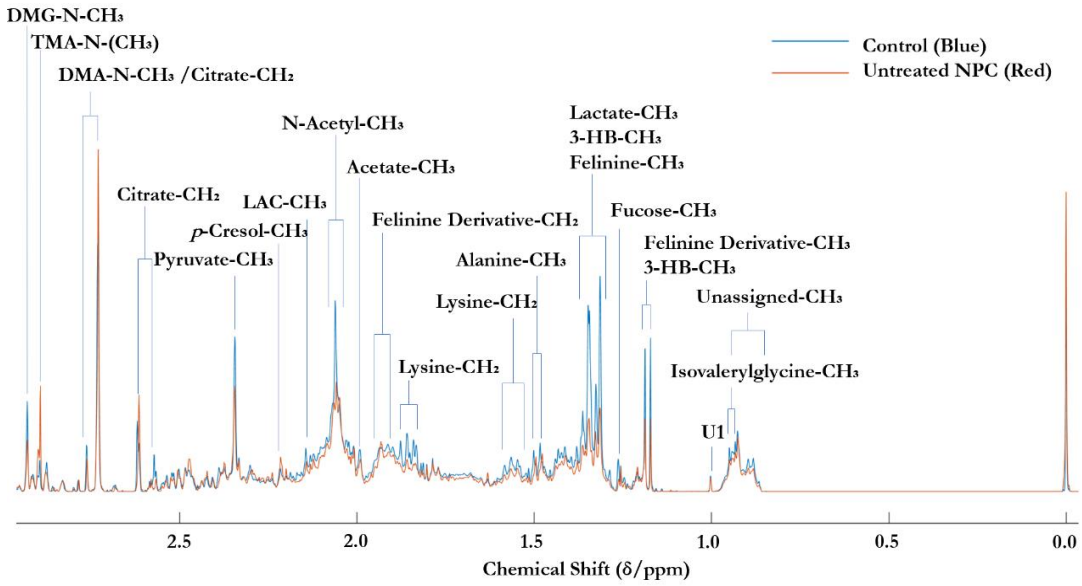
4.5 Data Normalisation, Scaling and Transformation

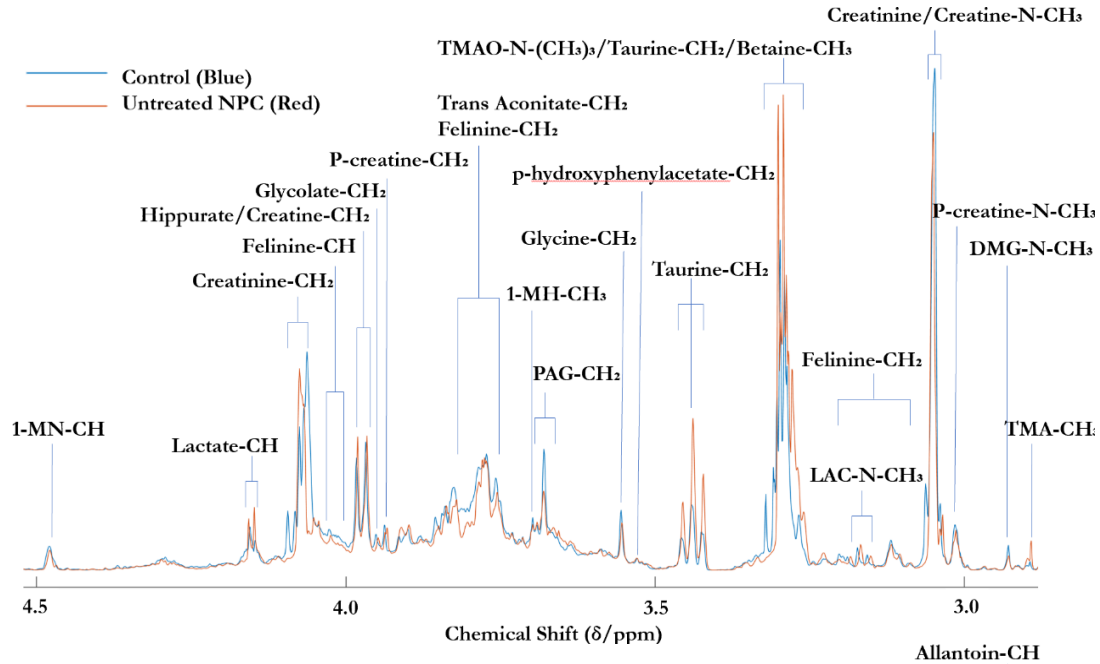
Raw integrals were constant sum normalised (CSN), Pareto-scaled and cube-root transformed prior to statistical analysis using *MetaboAnalyst* 4.0 [211]. CSN has been found to be recommended for urinary metabolomics datasets; other recommended normalisation strategies include probabilistic quotient normalisation [212].

4.6 Results

4.6.1 400 MHz ^{1}H NMR Urinary Profiles Acquired Using *noesygppr1d*

The 400 MHz single-pulse ^{1}H NMR spectra of these biofluids contained many prominent, sharp signals assignable to wide range of low molecular-mass biomolecules (Figures 4.1 and 4.2). Indeed, more than 80 metabolites were detectable in the spectra acquired on feline urine, and these included a range of biomolecules such as short-chain organic acid anions (e.g., acetate, formate, lactate, pyruvate and 3-D-hydroxybutrate); amino acids, including glycine, alanine, glutamate, glutamine, lysine, taurine, valine etc., together with N-acetylamino acids; carbohydrates, most especially NPC disease-relevant N-acetylsugars; and also nicotinate and nicotinamide pathway metabolites, amongst others. In addition to the amino acids above, the β-amino acid felinine was also assigned, which is exclusively found in feline urine; confirmation of the identity of resonances for this feline-specific metabolite is shown in the 2D NMR profiles of feline urine (Appendix 10).


DMG-N-CH₃
TMA-N-(CH₃)
DMA-N-CH₃ /Citrate-CH₂
N-Acetyl-CH₃
Acetate-CH₃
Citrate-CH₂
LAC-CH₃
p-Cresol-CH₃
Pyruvate-CH₃
Felinine Derivative-CH₂
Lactate-CH₃
3-HB-CH₃
Felinine-CH₃
Fucose-CH₃
Felinine Derivative-CH₃
3-HB-CH₃
Alanine-CH₃
Lysine-CH₂
Lysine-CH₂
Unassigned-CH₃
Isovalerylglycine-CH₃
U1
Control (Blue)
Untreated NPC (Red)
2.5
2.0
1.5
1.0
0.5
0.0
Chemical Shift (δ/ppm)


Creatinine/Creatine-N-CH₃
TMAO-N-(CH₃)₃/Taurine-CH₂/Betaine-CH₃
Control (Blue)
Untreated NPC (Red)
Trans Aconitate-CH₂
Felinine-CH₂
P-creatine-CH₂
p-hydroxyphenylacetate-CH₂
Glycolate-CH₂
Hippurate/Creatine-CH₂
P-creatine-N-CH₃
Felinine-CH
Glycine-CH₂
DMG-N-CH₃
Taurine-CH₂
Creatinine-CH₂
1-MH-CH₃
Felinine-CH₂
PAG-CH₂
1-MN-CH
Lactate-CH
TMA-CH
LAC-N-CH₃
4.5
4.0
3.5
3.0
Chemical Shift (δ/ppm)

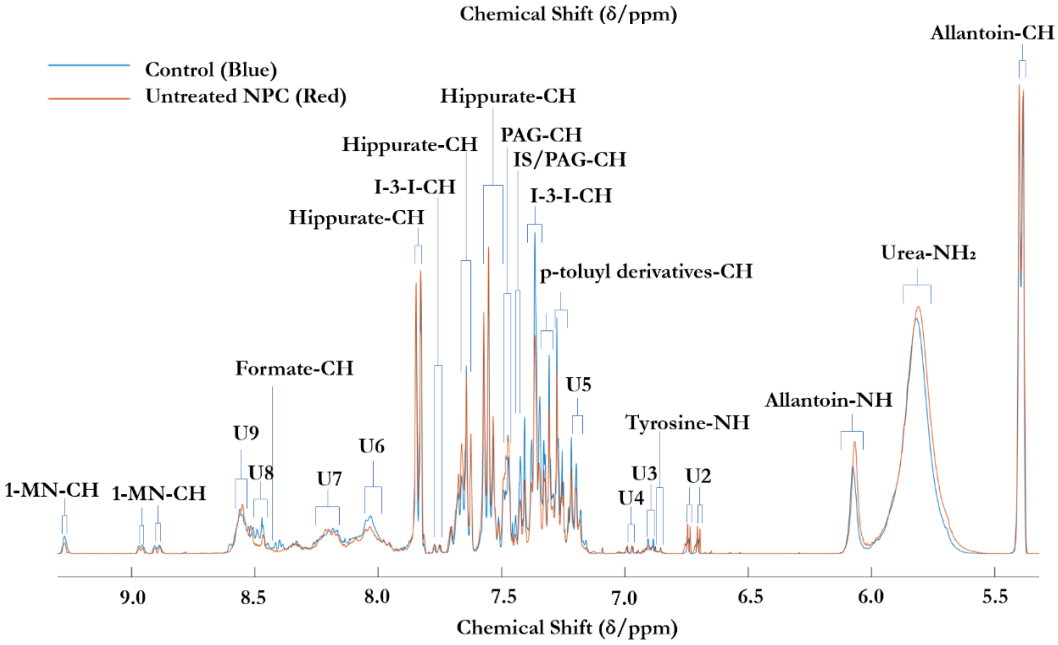

Allantoin-CH
Control (Blue)
Untreated NPC (Red)
Hippurate-CH
Hippurate-CH
PAG-CH
IS/PAG-CH
I-3-I-CH
I-3-I-CH
Hippurate-CH
p-toluyl derivatives-CH
Urea-NH₂
Formate-CH
U5
Allantoin-NH
U9
Tyrosine-NH
U6
U8
U7
U3
U2
U4
1-MN-CH
1-MN-CH
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
Chemical Shift (δ/ppm)

Figure 4.1. Average stack plots of the ^{1}H NMR profiles of NPC and control felines in red and blue, respectively, from 0.00 - 9.50 ppm. U1-U9 represent unknown metabolites, with the broad U6-U9 resonances presumably representing those of protein aromatic amino acids. Abbreviations: 1-MH, 1-methylhisitidine; 1-MN, 1-methylnicotinamide; 3-HB, 3-hydroxybutyrate; DMA, dimethylamine; DMG, dimethylglycine; I-3-L, indole-3-lactate; IS, Indoxyl sulphate; LAC, L-acetylcarnitine; NAC, N-acetyl groups; PAG, phenylacetylglycine; p-creatine, phosphocreatine; TMA and TMAO, trimethylamine- and trimethylamine-N-oxide-N($C\underline{H}_3$)$_3$ respectively.

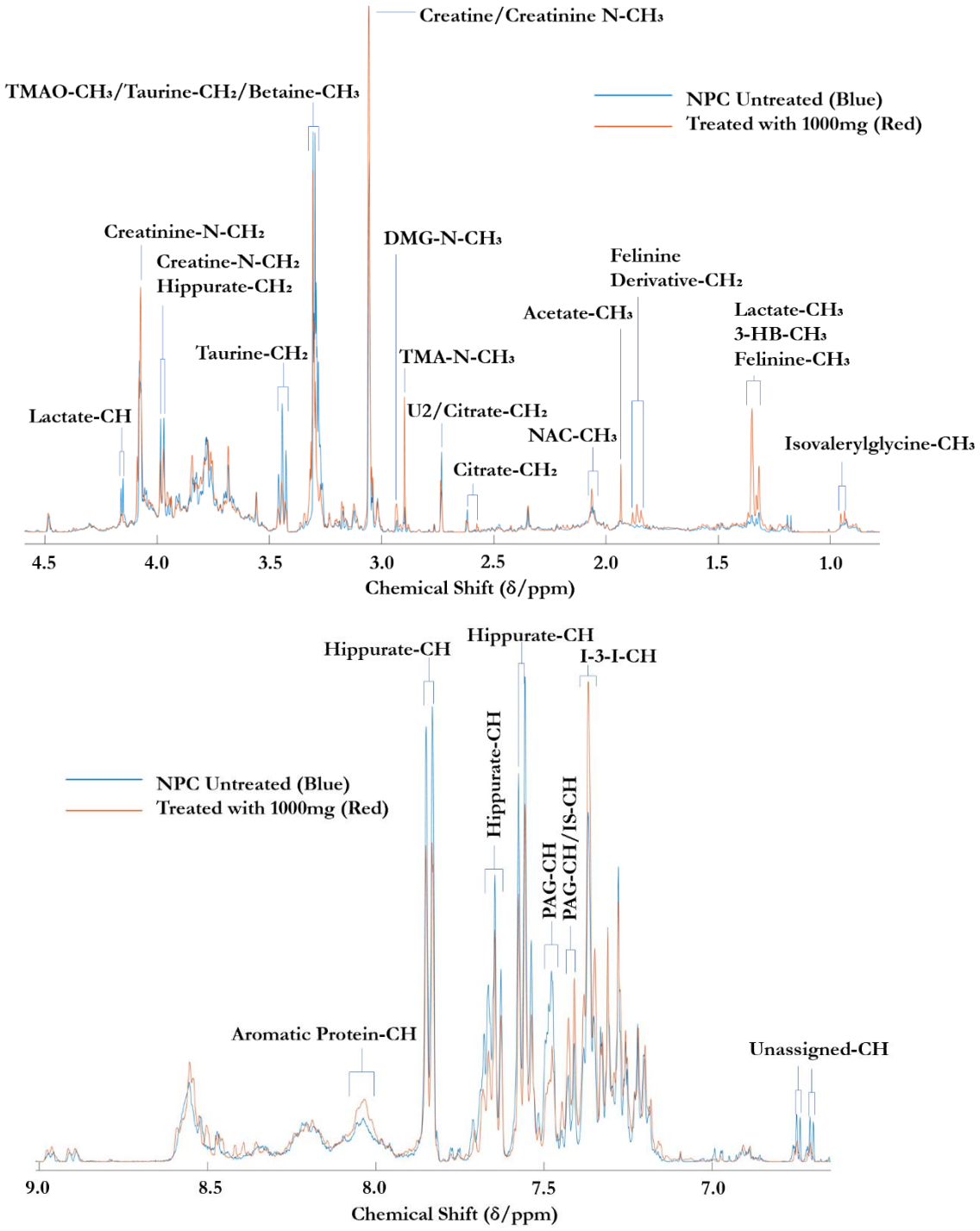


Figure 4.2. Average ^{1}H NMR profiles of NPC felines *versus* those treated with 1,000 mg HPβCD. Metabolite differences are assigned only. Abbreviations: 3-HB, 3-D-hydroxybutyrate; DMG, dimethylglycine; IS, indoxyl sulphate; I-3-L, indole-3-lactate; NAC, N-acetyl groups; PAG, phenyl-acetyl-glycine.

Comparison of different age and sex groups between control and NPC felines

For the first time, the age and gender of felines have been considered as part of a metabolomics investigation.

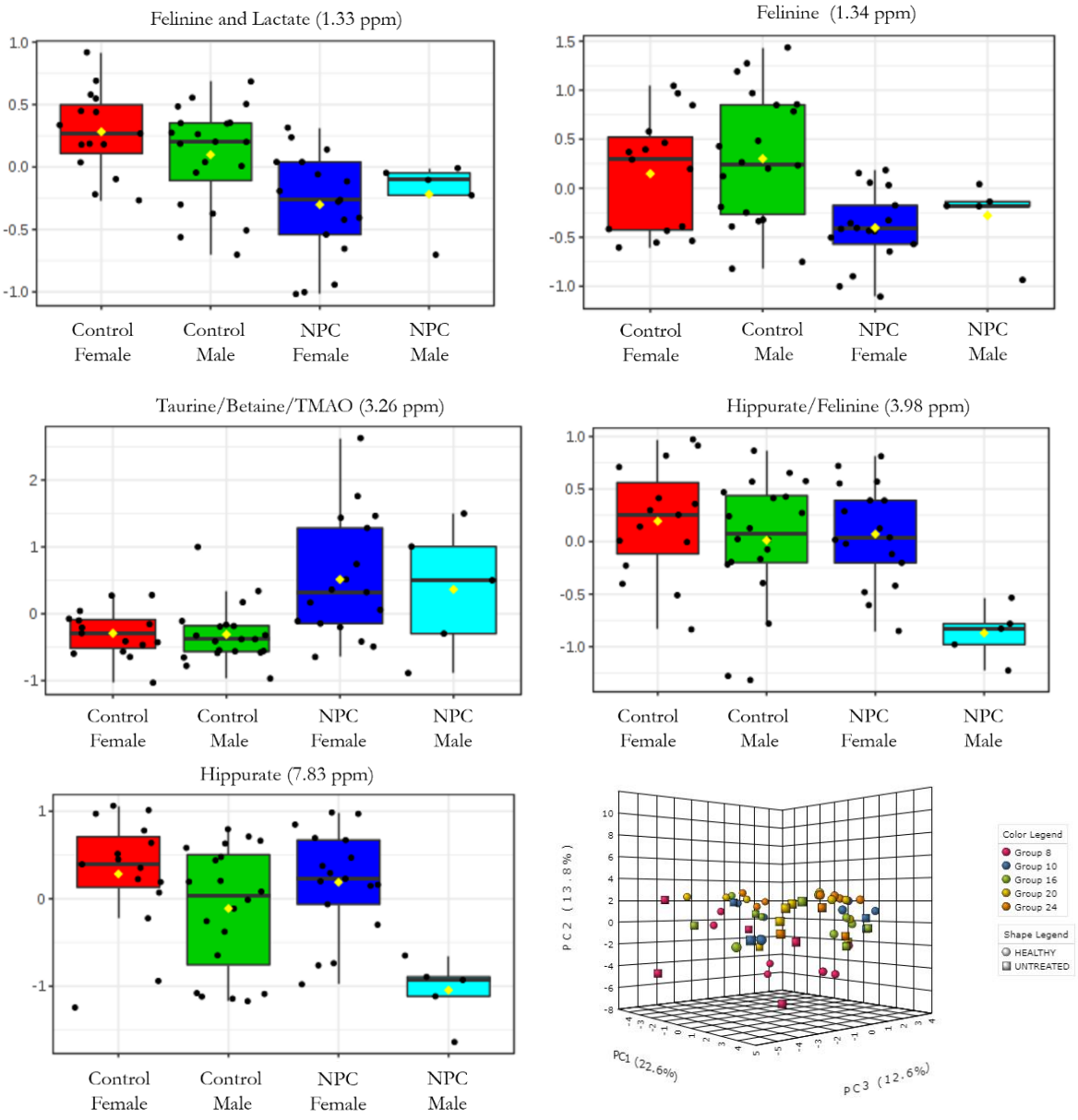


Figure 4.3. 8 significant bucket regions found by ANOVA involving comparisons of gender and disease status. Shown are mean±SEM plots for hippurate, taurine/betaine and felinine/lactate, which were found to be statistically significant ($p < 0.05$) This includes all datasets from 8–24 weeks for both control and NPC feline urine samples. Bottom right inset: Time series two factor analysis PCA scores plot of control *vs.* untreated NPC Felines at 8,

10, 16. 20 and 24 weeks. PC1, PC2 and PC3 represent 22.6, 13.8 and 12.6% of the total data variation respectively.

A univariate analysis of variance (ANOVA) approach was initiated on investigation of the feline urine in the control and NPC groups. Both Fishers and Tukey's test using parametric ANOVA showed 8 metabolites to be significant (Figure 4.3). The Kruskal Wallis test was also employed, and false discovery rate (FDR)-corrected statistic values were considered using a *p*-value of 0.05. These showed no significant metabolite variables. Upon employment of parametric ANOVA, no bucket regions were revealed to be significant. However, on separation of disease status and the age 8 weeks bucket regions, significant differences were found, and these may be related to growth and hormonal changes. These included buckets pertaining to hippurate, felinine and betaine/taurine (Figure 4.3). That containing combined taurine, betaine and TMAO signals are clearly elevated in the NPC group over those of the controls. However, the taurine δ = 3.43 ppm triplet resonance did not appear as significant in this test. This therefore suggests that it may not be taurine that is giving rise to this significant difference, but the other two metabolites (betaine and TMAO). Betaine has another resonance located at δ = 3.89 ppm, but this is in a relatively crowded region of the spectra and of a much lower intensity, and therefore may not appear as being statistically significant. TMAO only has one resonance (δ = 3.25 ppm), which is incorporated within this bucket region. 2D COSY analysis revealed no other correlations other than that with taurine between the signals in this bucket region.

All of the buckets incorporating hippurate resonances (δ = 3.95, 7.54, 7.62 and 7.82 ppm) appeared as significant buckets, and the same pattern was exhibited throughout, i.e. hippurate was elevated in controls and female NPC felines, and depleted in male NPC ones. Felinine was observed to be elevated in controls *versus* NPC disease across all bucket regions, with the exception of that with the neighbouring hippurate signal, i.e. NPC females were observed to have higher felinine levels. This can be attributable to the overlapping hippurate signal in this bucket region. The signal attributable to felinine at δ = 1.33 ppm shares the same bucket region as that of the lactate-CH_3 signal. However, the felinine triplet was divided across two buckets in order to overcome this problem, and facilitate the identification of felinine as the contributing factor for this significance, rather than that for lactate-CH_3. Additionally, the other lactate signal located at δ = 4.13 ppm did not show significance using this test, and this provided evidence that the felinine signal was the significant metabolite involved.

Principal component analysis (PCA) was performed considering two variables, the age of the feline upon sample collection, and the felines disease status, either control or NPC disease

(Figure 4.3, bottom right inset). PCA enables the outliers in this investigation to be removed for further statistical analysis, of which it appears that the younger felines are separating from the older ones, since the former group appear to be scoring more negatively on PC3. This PCA plot shows differences within age groups and between control and untreated felines, an observation suggesting that distinctions within the dataset are not only attributable to disease status, but also to age. Therefore, further investigation regarding which metabolites were affected by age and disease status were performed. Data showing differences between age-related metabolites is displayed in Appendix 9 in the form of heatmaps and overlaid average spectral profiles.

Multivariate analysis showed no significant differences between the four groups investigated, i.e. male controls, female controls, male NPC disease and female NPC disease. Pre-symptomatic felines were analysed at < 16 weeks of age. ANOVA showed no significant metabolites using the same testing criteria as above. This indicates that the severity of the disease is not also distinguishable by gender, and the changes observed in hippurate and felinine noted above are ascribable to disease status only.

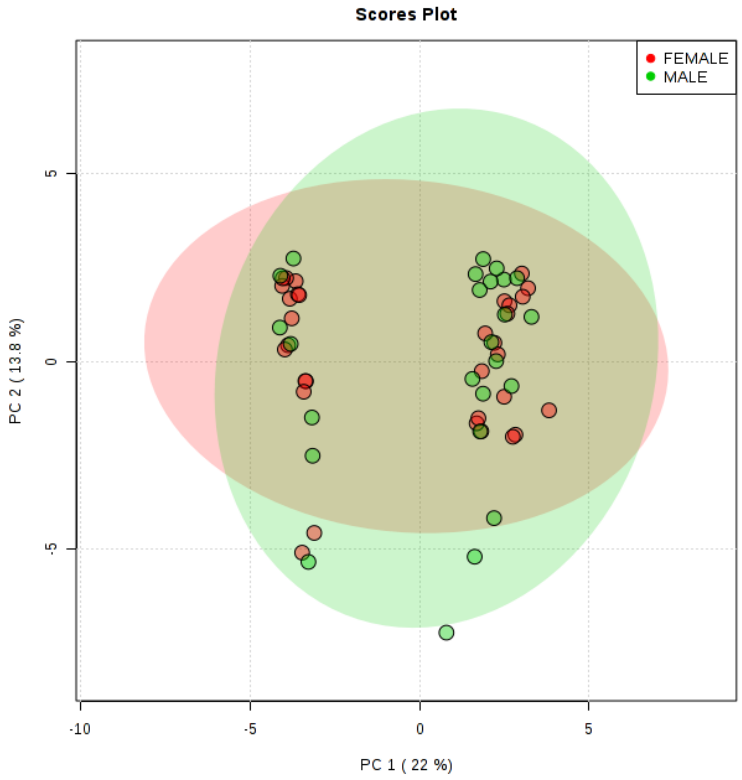


Figure 4.4. PCA scores plot of control felines (n = 35) *vs.* untreated NPC felines (n = 22) showing PC1 and PC2 representing 22.0 and 13.8% of the total dataset variance respectively from 8 - 24 age weeks, showing differences only between males and females. 95% confidence ellipses are shown for each group.

Figure 4.4 shows a comparison of males with females in both the NPC and control groups. There is only one outlier in this dataset, a male loading just outside the 95% confidence ellipse range. Two clusters are present, and this is attributable to disease status, since each individual point was checked to observe which scoring values were representing each sample. This demonstrates that the disease is not dependent on the sex of the feline, even though there is some variabilities between sexes.

Partial least squares- discriminant analysis (PLS-DA) cross-validation of the model showed a poor model fit and poor measure of consistency on comparison of genders (data not shown). Two sub clusters were observed in the data which were composed of a combination of disease and controls however, the controls were more prominent in one cluster and the disease was more prominent in the other cluster. However, as the model was invalid further testing had to be performed in order to distinguish whether the variability was due to disease or sex. So, a further strategy was performed with gender-only clusterings (Appendix 9).

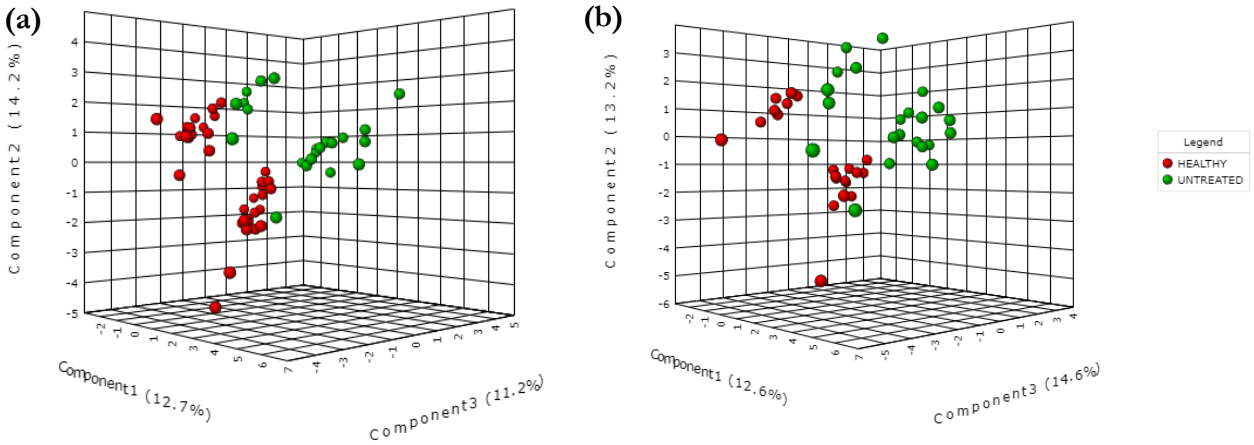


Figure 4.5. (a) PLS-DA plot of control felines (red) (n = 35) and untreated NPC felines (green) (n = 22) showing PC1, PC2 and PC3 representing 12.7, 14.2 and 11.2% of the dataset variance respectively. $Q^2 = 0.49$ when applying 10-fold cross validation methodology using a 6 component model and $p = 0.084$ was determined using prediction accuracy during training as the permutation test, using 2000 permutations **(b)** PLS-DA plot of control felines (n = 21) vs untreated NPC felines (n = 22) showing PC1, PC2 and PC3 representing 12.6, 13.2 and 14.6% of the dataset variance respectively. $Q^2 = 0.62$ using a 6 component model applying 10-fold cross validation methodology and $p = 0.0125$ was determined using prediction accuracy during training as the permutation test, using 2000 permutations.

PLS-DA has shown that the urine profiles are insignificant comparing control felines compared to NPC (Figure 4.5, (a)). Due to the disease being a neurological disorder and progressing further over time, the older felines with NPC disorder, cluster closer to the younger controls. Therefore, a ^{1}H NMR methodology when using an age-matched feline model, might not be the best approach for early onset disease diagnostics suggesting early biomarkers for NPC. However, if the younger control felines are removed from this model

(weeks 8-10) the data becomes significant and the model becomes accurate hence, this is a better model for observing which metabolites are causing the changes in the later manifestations of NPC (Figure 4.5, (b)). Felines typically become fertile around 16 weeks of age and therefore, the extra variability of fertile and unfertile felines and the hormonal changes associated with this difference is removed. There are two clear sub clusters in the dataset which were investigated. It was found that these sub clusters were not due to age or sex (Figure 4.5, (b)). Bucket regions that are causing a significant difference in the PLS-DA plot were assigned and it is also defined as upregulated or downregulated in disease profiles respectively (Table 4.1).

Significant Bucket (δ/ppm)	**Upregulated / Downregulated in NPC Disease**	**Assignment**
1.33-1.35	↓	α-Hydroxyisobutyrate-CH_3 (*d*)/Felinine-CH_3 (*d*)/Lactate-CH_3 (*d*)
3.25-3.27	↑	Taurine-CH_2-NH_3^+ (*t*)/Betaine-CH_3 (*s*) /TMAO-CH_3 (*s*)
3.41-3.47	↑	Taurine-CH_2-SO_3- (t)
3.03-3.07	↓	Creatinine-CH_3 (*s*)/Creatine-CH_3 (*s*)
1.31-1.32	↓	α-Hydroxyisobutyrate-CH_3 (*d*)/Felinine-CH_3 (*d*)/Lactate-CH_3 (*d*)
3.64-3.70	↑	PAG-CH_2 (*s*)
1.32-1.33	↓	α-Hydroxyisobutyrate-CH_3/Felinine-CH_3/Lactate-CH_3 (*d*)
1.85-1.87	↓	Lysine-CH_2 (*m*) /Felinine Derivative-CH_2 (*m*)
3.33-3.36	↓	Felinine-CH_2 (*m*)
2.90-2.94	↓	Dimethylglycine-CH_3 (*s*)
1.82-1.85	↓	Felinine derivative-CH_2 (*m*)
3.32-3.33	↓	Felinine-CH_2 (*m*)
1.35-1.37	↓	Felinine-CH_3 (*d*)
1.24-1.27	↓	L-Fucose-CH_3 (*d*)
2.75-2.77	↓	Citrate-CH_2 (*d*)/DMA-$N(CH_3)_2$ (*s*)

Table 4.1. Top 15 significant metabolites assigned that are upregulated or downregulated in NPC felines from the PLS-DA strategy applied (Figure 4.5, (b)). Abbreviations: PAG, phenyl-acetyl-glycine. Multiplicities are indicated in brackets: multiplet, triplet, doublet and singlet are denoted by (*m*), (*t*), (*d*) and (*s*) respectively.

Metabolites that appear as significant include taurine, TMAO, betaine, felinine which has been previously linked to being significant in Figure 4.3. Hippurate was not highlighted as one of the top 15 significant metabolites unlike the previous test in Figure 4.3. Other metabolites that were shown to be significant include buckets that are attributable to creatinine/creatine, fucose, citrate and dimethylglycine. This is a representation of the importance of multivariate testing and univariate testing to distinguish different metabolites as significant, as some metabolites can be masked due to the quantity of bucket regions.

ANOVA (Figure 4.6) and fishers LSD/Tukeys test was repeated with the 16-20+ week datasets and both tests provided the same significant bucket regions. Some metabolites appeared significant that previously had not during MV analysis, this included bucket regions attributable to pyruvate and DMA/glutamine. Felinine and taurine bucket regions appeared as significant but hippurate was not a significant metabolite in this testing.

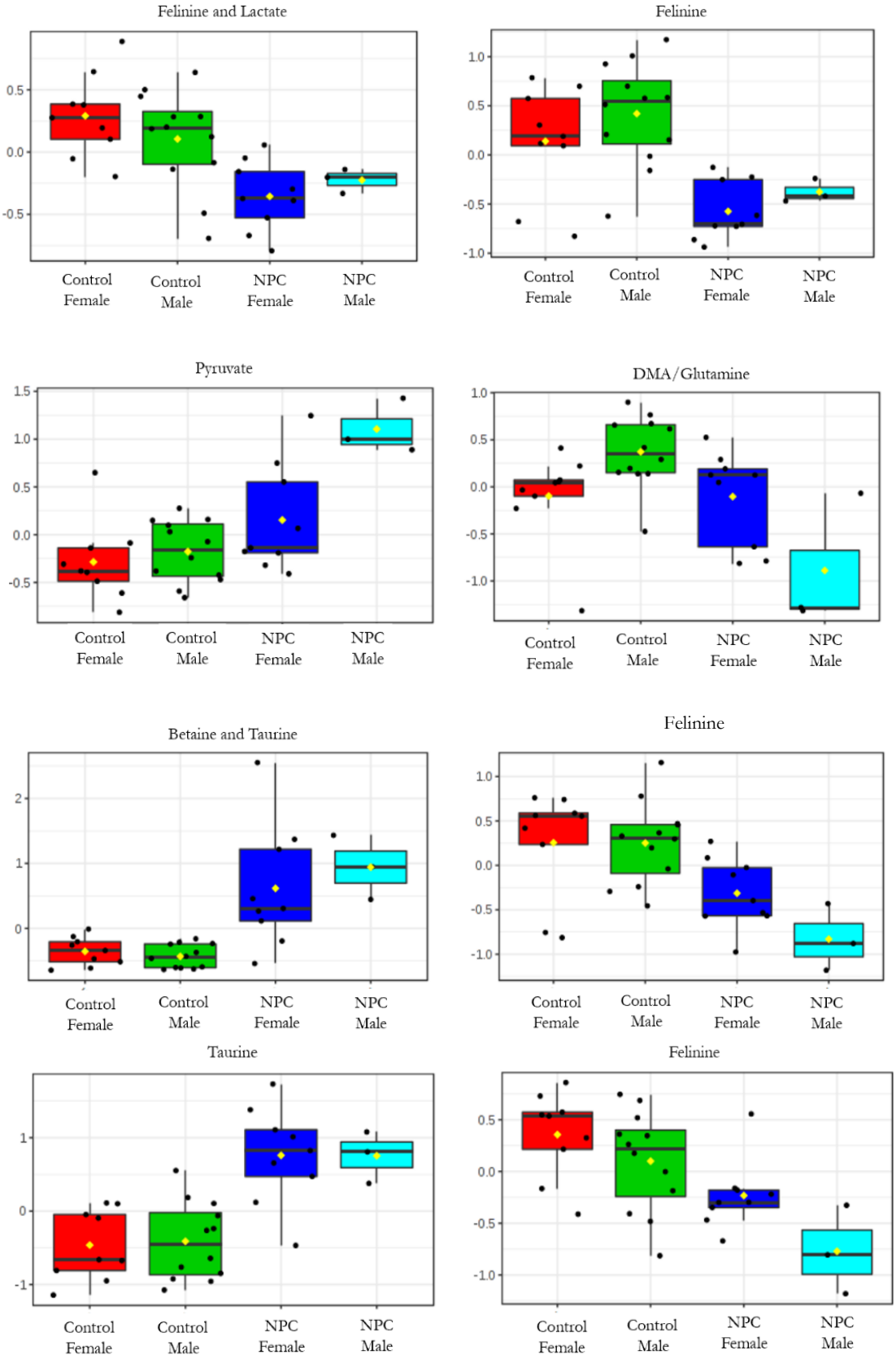


Figure 4.6. ANOVA analysis of control *versus* NPC felines grouped by gender, showing data from 16 – 24 weeks only. Above are 7 of the metabolites which showed statistical significance. Mean±SEM plots show those for pyruvate, taurine/betaine/TMAO, DMA/glutamine and felinine.

Further statistical analysis was performed in order to distinguish between the separation observed between the two cohorts. The sub-clustering of the controls and untreated NPC felines were explored and hence felines were grouped into female and male groups in order to investigate this further (Figure 4.7). However, from PCA and PLS-DA approaches it can be observed that this is indeed not the cause of the separation; therefore, it appears that this division arises between the datasets arises from another source. However, these models may be overfitted in view of the small number of samples in each group and the permutation testing value was not significant, and PLS-DA usually requires larger sample sizes per group.

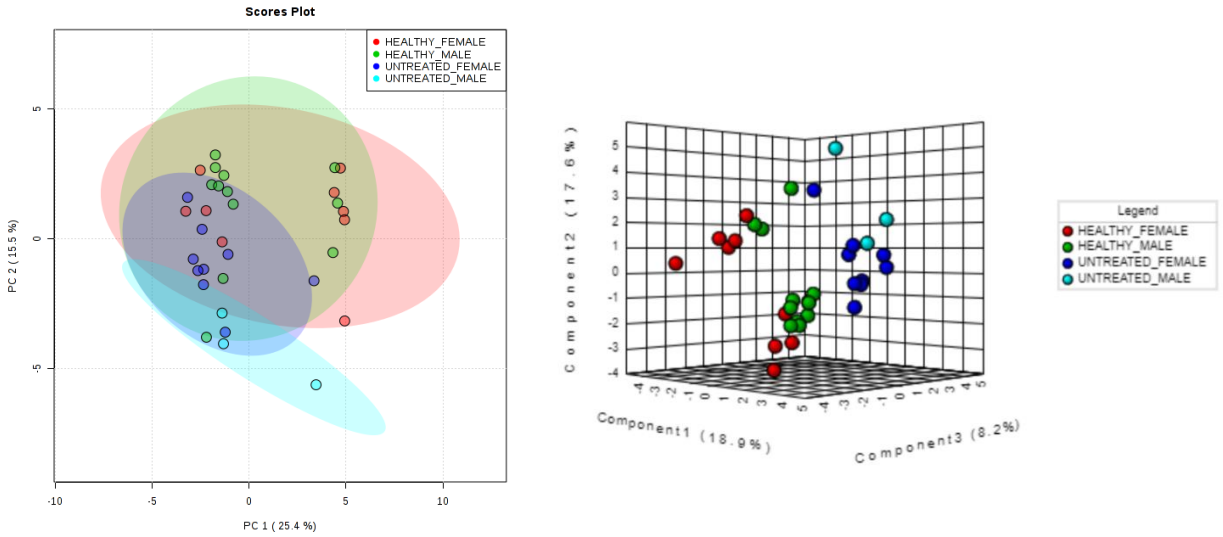


Figure 4.7. (a) PCA scores plot of control females, control males, untreated females and untreated males in red, green, blue and light blue respectively. PC1 and PC2 represent 25.4 and 15.5% of the total dataset variance respectively. 95% confidence ellipses are mapped, and there are no evident outliers in this dataset. **(b)** PLS-DA plot of control females, control males, untreated females and untreated males in red, green, blue and light blue respectively. PC1, PC2 and PC3 represent 18.9, 17.6 and 15.5% of the total dataset variance respectively.

4.7 Metabolic Differences between NPC felines and felines treated with HPβCD

Three treatment regimens were trialled. In the first group, NPC disease felines were treated with Cyclodextrin/Kleptose IT (120 mg) early, i.e. initiated at 3 weeks of age (n = 36), whilst the second received the same treatment regimen but initiated at 16 weeks of age (n = 36). The third group of felines was treated with a combination SQ (1,000 mg/kg) and IT (120 mg) early (n = 23).

Student's t-tests of all the bucket regions showed no significant differences between metabolite levels on comparison of Cyclodextrin/Kleptose IT 120 mg early group *versus* the

untreated NPC feline one. PCA showed no separation between these groups. PLS-DA model of NPC felines treated with Cyclodextrin/Kleptose IT 120 mg early-initiated at 3 weeks of age *versus* untreated NPC felines was accurate with the highest $Q^2 = 0.53$ on a model containing 8 components, but showed no significance ($p > 0.05$) determined from permutation testing, which invalidated the model. Machine learning methods were initiated, but were also unable to distinguish between these two groups.

Student's t-tests showed no significant metabolites on comparison of Cyclodextrin/Kleptose IT 120 mg late group *versus* the untreated NPC feline one. PCA showed no separation between these groups. A PLS-DA model of NPC felines treated with Cyclodextrin/Kleptose IT 120 mg late *versus* untreated NPC ones showed no clustering, with the highest Q^2 value of only 0.24 achieved on a 6 component model; similarly, no significance was found on permutation testing ($p > 0.05$). Machine learning methods were again unable to distinguish between these two groups.

However, the third group was treated with a combination SQ (1,000 mg/kg) and IT 120 mg early showed separation between from untreated NPC felines using PCA, and also significance and accuracy in PLS-DA assessments (Figure 4.8). Three outliers were noted in the PCA approach.

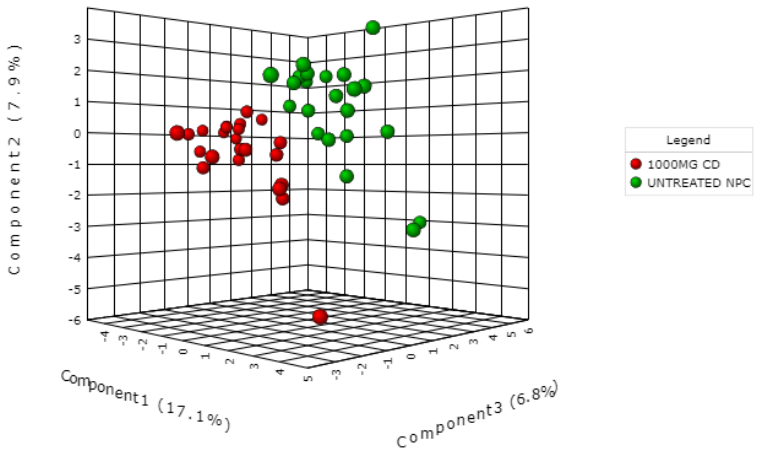


Figure 4.8. PLS-DA scores plot of untreated NPC felines (n = 22) *vs.* cyclodextrin-treated felines (a combination of SQ (1,000 mg/kg) and IT 120 mg early was administered) (n = 23), showing that PC1, PC2 and PC3 representing 17.1, 7.9 and 6.8% of the total dataset variation respectively. $Q^2 = 0.53$ when applying 10-fold cross validation methodology on a 4 component model, and $p = 0.028$ determined using prediction accuracy during training as the permutation testing regimen, which used 2,000 permutations.

Significant Bucket (δ/ppm)	Upregulated / Downregulated in treated NPC felines	Assignment
1.33-1.37	↑	α-Hydroxyisobutyrate-CH_3 (*d*)/Felinine-CH_3 (*d*)
1.30-1.33	↑	α-Hydroxyisobutyrate-CH_3 (*d*)/Felinine-CH_3 (*d*)/Lactate -CH_3 (*d*)
3.03-3.07	↑	Creatine-CH_2 (*s*)/Creatinine-CH_2 (*s*)
1.82-1.88	↑	Felinine Derivative-CH_2 (*m*)
3.41-3.46	↓	Taurine-CH_2-SO_3- (*t*)
0.86-0.87	↑	Unassigned-CH_3 (*broad*)
3.25-3.27	↓	Taurine-CH_2-NH_3^+ (*t*) /Trimethylamine-N-oxide-CH_3 (*s*)
3.31-3.35	↑	Unassigned (*broad*)
7.67-7.71	↓	Hippurate-CH (*t*)
4.13-4.18	↓	Lactate-CH (*q*)
2.40-2.41	↓	Glutamine-CH_2 (*m*)
4.05-4.10	↑	Creatinine-CH_2 (*s*)
3.64-3.70	↓	PAG-CH_2 (*m*)
3.27-3.31	↓	Taurine-CH_2-NH_3^+(*t*) /Trimethylamine-N-oxide-CH_3 (*s*) /Betaine-CH_3 (*s*)
8.47-8.49	↓	Tentative Protein Aromatic Amino Acids-CH (*broad*)

Table 4.2. Top 15 significant bucket regions with metabolites assigned that were upregulated (↑) or downregulated (↓) in treated NPC felines using the VIP scoring system. Abbreviations: PAG, phenyl-acetyl-glycine. Multiplicities are indicated in brackets: multiplet, triplet, doublet and singlet and denoted by (*m*), (*t*), (*d*) and (*s*) respectively.

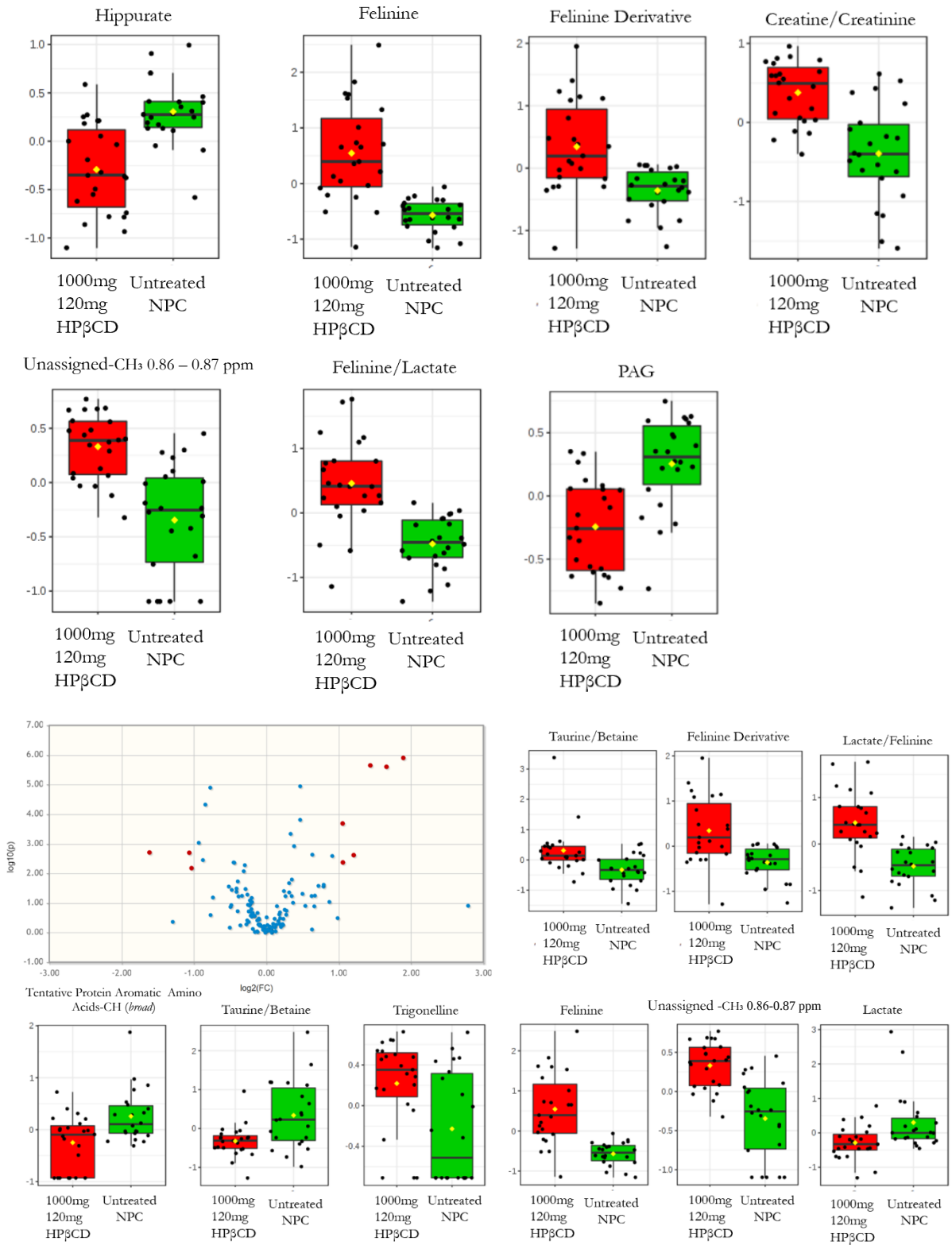


Figure 4.9. (a) Student's t-tests using equivalent group variances with an adjusted *p* value cut off value of 0.005. Bucket regions are labelled with corresponding assignments and include hippurate, felinine, lactate, phenylacetylglycine, creatine/creatinine and a felinine derivative. **(b)** Volcano Plot showing significance (y-axis) *versus* fold-change (x-axis) in which the fold threshold is 2, and the equivalent group variance *p* value threshold is 0.1. Bucket regions in red are significant, and those in blue are insignificant. Bucket regions values were

then displayed as box and whisker plots, in which red and green show the 1,000 mg/kg SQ and 120 mg IT HPβCD and untreated NPC groups, respectively.

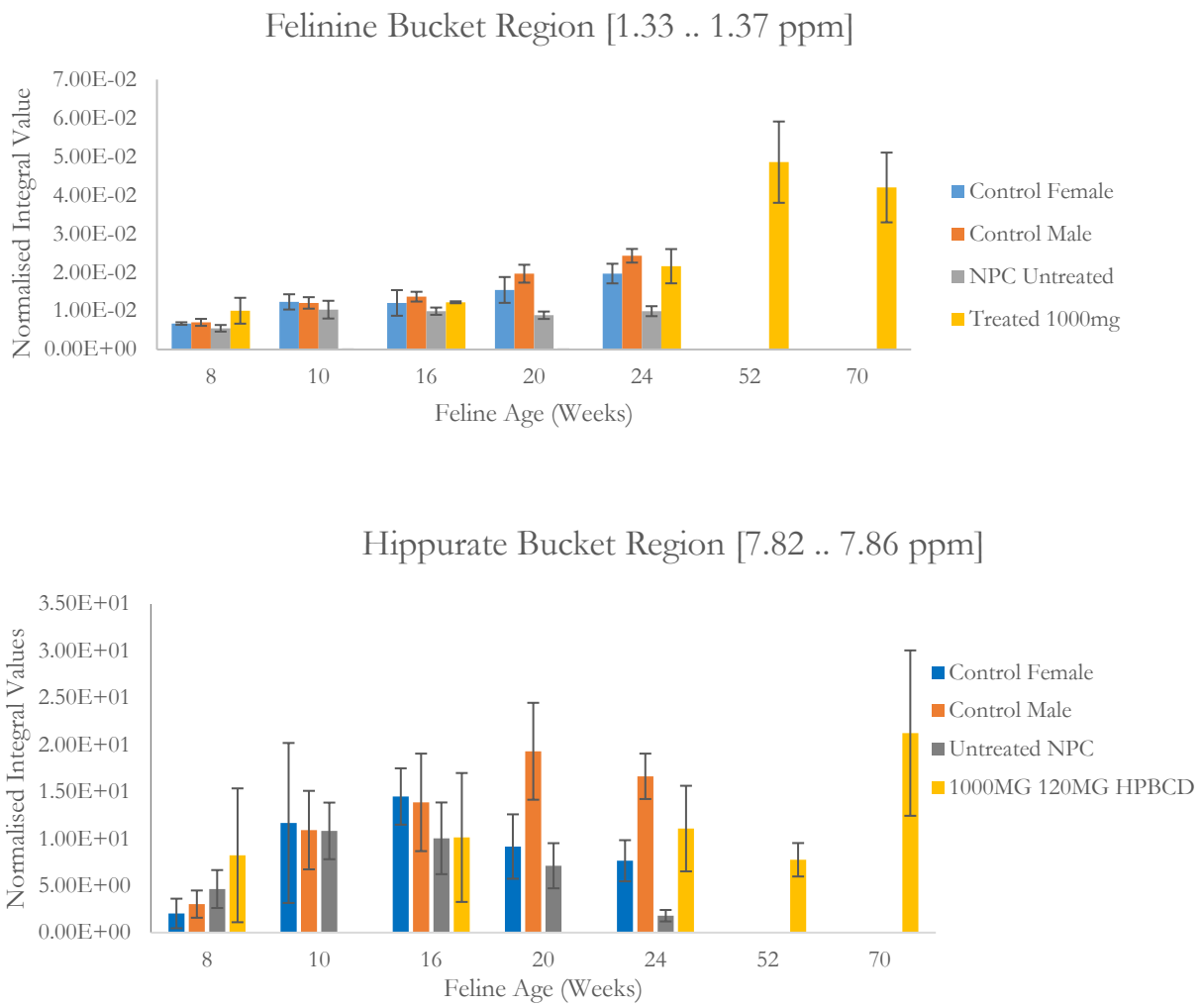


Figure 4.10. Mean±SD TSP-normalised integral values of felinine and hippurate bucket regions with corresponding feline age and treatment type. For the control group, males and females have been divided in order to evaluate between-gender differences. Data was not collected for treated 1,000 mg/kg SQ and 120 mg IT HPβCD felines at 10 weeks of age.

4.8 Discussion

Felines have been previously shown to be ideal models for observing pathological changes in NPC disease for selected studies, such as those involving cholesterol-restricted diets [213] and treatment outcomes [204]. However, the feline urine is yet to be investigated in depth, inclusive of how feline metabolism changes with age and gender. Moreover, how these changes are affected by a disease, such as NPC, and how treatment impacts on urinary metabolism has yet to be explored in detail. This study, for the first time, has observed such urinary metabolic changes over time, along with those arising from the application of treatment regimens. This work could underpin new areas of metabolomics investigations into feline urine and support experimental design considerations. Some urinary profiles showed metabolites which were unassigned in this study, and were excluded (Appendix 7).

Felinine and Felinine Derivatives

Present research suggests that felinine is a testosterone-dependent urinary metabolite, and that this feline-specific β-amino acid may be involved in territorial marking [214], [215]. Injections of testosterone increased felinine levels, and that of its derivatives, in feline urinary profiles in male and female cats [215]. Interestingly, felinine is involved in two pathways and can be synthesised in the liver or kidney [216]. Both pathways involve the precursor dimethylallyl-PP (isopentyl-PP), which is involved in the biosynthesis of cholesterol and steroids [216]. Felinine in the 1D NMR spectra was determined using a 2D TOCSY analysis strategy (Appendix 10). Results from the ^{1}H NMR urinary profiles acquired here demonstrated that felinine concentrations are higher in controls and treated 1,000 mg/kg SQ and 120 mg IT HPβCD felines than those of the untreated NPC group. Additionally, results observed in plasma show cholesterol is elevated in untreated NPC felines, and depleted in 1,000 mg/kg SQ and 120 mg IT HPβCD-treated and control felines (Appendix 12). NPC felinine is depleted, and this observation suggests a possible promotion of the steroid and cholesterol synthesis pathways. This can be readily confirmed from the average spectral profiles (Figures 4.1 and 4.2), mean±SEM levels (Figure 4.10), and by univariate (Figures 4.3, 4.6 and 4.9) and multivariate statistical analysis strategies (Figures 4.3-4.5, 4.7 and 4.8).

In NPC murine models, overall testosterone secretions are decreased, and low serum testosterone levels were observed in NPC disease when compared to those of healthy controls [217]. Low testosterone alone cannot serve as a marker for NPC disease, since it is found in other phenotypes such as stress, anxiety, ageing and infertility. Nevertheless, this supports findings the findings made here, i.e. felinine has been shown to decrease in urinary

profiles in untreated NPC *versus* control felines. Felinine increases again upon treatment with 1,000 mg SQ and 120 mg IT HPβCD, and this possibly suggests a positive therapeutic outcome. However, these changes were not observed in the 120 mg-dosed felines treated both pre and post-symptomatically. The same pattern is observed with the felinine derivative. This felinine derivative has been assigned previously [218] and signals can be observed at δ = 1.20 and 1.80 ppm.

Age has also been shown to impact on the rate of felinine production in felines, and this was observed from mean±SEM level plots (Figure 4.10), and also in the ^{1}H NMR spectra acquired (Appendix 9). The SEM error bars are relatively small, with very little overlap between the groups. Gender-related changes have also been observed for urinary felinine; specifically, males have higher levels than females (Figure 4.10). Indeed, felinine has been shown to be elevated in entire male cats *versus* entire female and castrated cats, both herein and elsewhere [214], [215]. Age-related modifications in felines have also been previously documented in the literature [219]. However, for the first time it has been observed that NPC felines have significantly lower levels of this metabolite, despite gender differences. The age-related differences in felinine are only observed in the NPC disease group post-symptomatically, i.e. past 16 weeks (Figure 4.10). Upon treatment of this condition with 1,000 mg SQ and 120 mg IT HPβCD, levels appear to return to those similar to the control group.

Taurine, Betaine and TMAO

In a previous human urinary study, TMAO and taurine were observed to increase in male participants, and decrease in females [220]. Furthermore, these metabolites were observed to increase in older participants compared to younger subjects in this human study [220]. Additionally, urinary taurine in rats was also found to increase with age [221]. Urinary taurine and TMAO have also been correlated with dietary changes, specifically the Atkins diet, which consists of high fat, high protein and low carbohydrates, and which is observed more frequently in the Swedish population than it is in the British [222]. TMAO has been observed to be highly variable in human subjects [222]. Nevertheless, there are notable correlations of upregulated TMAO with increased endothelial dysfunction, adhesion of monocytes through activation of PKC NF-kB VACM-1 pathways, pro-inflammatory cytokines *via* MAPK and NFkB, tumour necrosis factor, interleukin IL-1β, increased cardiovascular disease and reduced endothelial self-reparation [223]. TMAO has also been observed to reduce the expression of intestinal cholesterol transporter Niemann-Pick C1-like1 (which is

downregulated by high cholesterol diets), and urinary TMAO has also been demonstrated to be elevated in neurological disorders [223].

All felines were on the same diet and were of the same breed, therefore such changes observed here are not attributable to population or diet. Interestingly, taurine was noted to be markedly increased in the NPC felines over that of the controls, and was observed in the top 25 metabolites in the heatmap (Figures 4.8 and 4.9). This is of much importance, since this study is not age- and gender-matched, which is a vital consideration in the experimental design of metabolomics investigations. The felines were treated survive longer, and therefore it is to be expected that they have higher levels of TMAO and taurine. However, markedly higher levels of TMAO were observed in NPC felines; however, the TMAO resonance bucket region is unfortunately shared with signals arising from betaine and taurine. Elevated urinary TMAO in NPC disease, and reduced levels in controls and treated felines suggests that this variation could arise from altered cholesterol and sterol metabolism in NPC felines. This has also be documented elsewhere that TMAO has such an effect on cholesterol and sterol metabolism [224]–[226]. Disturbances in methylamine metabolism have also been discussed in previous works [13], and elevations in methylamines have been observed in NPC disease urine [13]. The findings would support this previous hypothesis, and also suggest a disturbance in gut microflora in NPC disease [13].

Urinary taurine has been suggested previously as a non-invasive biomarker of liver damage [227], and in NPC, liver damage is a symptom, in the form of splenohepatomegaly. A positive correlation between urinary taurine and serum AST has been previously observed [227], and this is consistent with the findings made here. Taurine was observed to be elevated in NPC felines *versus* controls in both univariate and multivariate testing strategies (Figure 4.7 and Table 4.1). Also, taurine was observed to deplete in SQ 1,000 mg/IT 120 mg-treated felines *versus* untreated NPC ones (Table 4.2).

Dimethylglycine

Dimethylglycine has appeared as a significant biomarker metabolite in NPC felines when compared to the control group using multivariate statistical analysis (Table 4.1). However, using multivariate analysis, a reduction in dimethylglycine was not observed on treatment (Table 4.2). This could possibly be attributable to the slightly different age ranges in felines, since urinary dimethylglycine has been observed to elevate with age (Appendix 9). The control felines have survived for a slightly longer period than the NPC disease felines, and this possibly indicates the significance of this metabolite. However, an elevation in

dimethylglycine rather than a depletion would be expected in view of disease-mediated disturbances in gut microbiota, since dimethylglycine is a product of choline degradation and has been shown to be modulated in diseases related to cardiovascular disease, obesity and diabetes [228].

Creatinine/Creatine

Creatinine is a breakdown product of creatine. Creatinine has been demonstrated to be elevated in human males over those of females [220]. Furthermore, creatinine was observed to be elevated in older human participants over that younger human subjects [220], and a similar trend was also observed in rats [221]. However, the inverse is correct for creatine in human participants [220]. The combined creatinine/creatine >N-CH_3 resonance (δ = 3.03-3.07 ppm) is depleted in 8-week control group when expressed relative to control 24-week one (Appendix 9). Interestingly, this signal does arise in the top 15 significant metabolites in NPC disease felines, suggesting an abnormal aging process, since it is depleted significantly in the controls (Table 4.1). Upon treatment with HPβCD, however, this creatinine/creatine signal intensity elevates (Table 4.2), and this was also shown to be significant in a univariate t-test (Figure 4.10). However, this could also be ascribable to the skewed male:female ratio of the NPC cohort (Table 4.0).

These results are particularly interesting, since in the plasma a lower creatine level of felines with SQ 1,000 mg/IT 120 mg treated and control groups was observed, together with an elevation in NPC and the other two pre- and post-symptomatic 120 mg treatments (Appendix 12). These results are the converse of those observed here for the ^{1}H NMR urinary profiles. Previous investigations have observed no changes in plasma creatinine in a mouse model monitoring NPC disease and the impact of HPβCD treatment on it [229], but this is not what was observed here in feline models. It could be hypothesised that creatine is lowered in the SQ 1,000 mg/IT 120 mg HPβCD-treated felines versus the untreated NPC ones in view of age-related differences between these groups. This is a most likely explanation, since urinary creatine has be previously been observed to be elevated in a human model [13].

Hippurate

In a previous study, urinary hippurate was noted to be downregulated in female urine with respect to that of males [220]. Furthermore, younger participants have lower hippurate than older participants in humans [220] and rats [221]. Hippurate does not appear as a significant metabolite on comparison of ageing control felines in the heatmaps with the top 25 most

significant metabolites shown (Appendix 9); however, NPC felines exhibited an elevation in hippurate upon aging rather than a decline (Appendix 9). Figure 4.10 demonstrates that hippurate is highly variable amongst all felines at a range of ages, including those with NPC disease. It is not until 24 weeks of age that hippurate was significantly decreased in NPC felines over those of all other groups (Figure 4.10). Urinary hippurate has been observed to be decreased in Crohn's disease, and has also been linked to alterations in gut microbiota [230]; this supports the findings made in this study.

PAG

PAG has been previously suggested as a putative biomarker of phospholipidosis [231], [232], another lysosomal storage disorder. In these particular studies, PAG was elevated in urinary profiles. From the results acquired here, PAG has been observed to be elevated in NPC disease over those of controls, and those treated with 1,000 mg SQ/120 mg IT HPβCD. PAG was also observed to deplete over time (8 *versus* 20 weeks) in control felines, whereas the converse was the case for NPC disease felines (Appendix 9). This metabolite is evidently age-related, but the pattern differed from those of the disease and control groups. This would suggest that PAG may serve as a valuable marker for the detection and treatment outcomes of NPC disease in felines. Similarly to the study performed on phospholipidosis [231], [232], citrate, a TCA intermediate, was also found to be depleted in NPC urine compared to control urine. In this previous investigation, it was proposed that mitochondrial citrate production for energy was found to be imbalanced [232]. PAG is part of phenylalanine metabolism, and disruption of this pathway has also been previously linked to mitochondrial disruption.

PAG has been shown to be depleted in pigs undergoing treatment from liver injury [233]. Although there is liver injury in NPC disease in the form of hepatosplenomegaly, these results do not correlate with the findings made in this investigation, since PAG is elevated in NPC disease and not downregulated. Hence, it is more likely that the decrease observed in PAG is attributable to mitochondrial toxicity in NPC felines, and such a disturbance may arise from this rather than liver injury.

Citrate

As noted above, citrate was observed to decrease in NPC and control felines on comparison of the 8-to 20-week datasets (Appendix 9). These age-related changes might help understand why NPC urinary citrate levels are downregulated. The citrate ^{1}H NMR bucket region (δ = 2.75-2.77 ppm) is, however, partially shared with another metabolite (DMA).

In a human model, citrate was decreased in males relative to females [220]. Furthermore, citrate was also observed to be elevated in older human subjects over those of younger ones [220]. However, upon monitoring citrate in rats from 4 – 20 weeks, citrate was seen to slowly decrease from 4 – 10 weeks and then stabilise at a 20 week duration [221]. Data collected in this study to be consistent with these reports.

Lactate

The bucket region attributable to lactate's -CH_3 function signal (δ = 1.33 – 1. 37 ppm) was observed to decrease in the urinary profile's comparison of untreated NPC *versus* controls. However, this bucket region is also shared with felinine's-CH_3 resonance. Lactate has another (-CH function) bucket region at 4.13 – 4.18 ppm, which was insignificant in this initial test. Lactate is elevated in female human urine over that of males [220]. Despite this, both buckets pertaining to lactate are depleted in NPC, despite the elevation of female felines monitored as part of this investigation. However, without further testing, it is difficult to distinguish whether lactate in urine is depleted in these samples in view of its overlap with another signal; this can be observed in the acquired 2D COSY spectroscopy (Appendix 4). The 2D spectral correlation of lactate resonances in the 1D NMR spectrum was confirmed using a 2D COSY approach (Appendix 10). This 2D profile shows several other correlations within this bucket region. Further work on the urinary profiles to complete the 2D spectral quantification of lactate could be performed in order to confirm the significance of this particular metabolite. Lactate has been previously observed to be upregulated in the presence of selected diseases, and as such this observation is unusual. Lactate is hypothesised to not be the cause of this bucket region to be significant, when comparing the difference between NPC and control groups, since it is generally elevated in disease states in other investigations [234], which is the converse of what was observed here.

However, on comparison of untreated NPC disease felines with those treated with SQ 1,000 mg/IT 120 mg HPβCD, the lactate bucket region 4.13 – 4.18 ppm is downregulated in the latter group. This is more consistent with what has previously observed in the literature, and elevated lactate levels may have some potential to be used as a biomarker in NPC disease.

Outcomes in Serum, Liver and Brain Biochemistry

Prior to investigating the urinary profiles in this study, it is important to consider the previous findings of investigations focused on serum, liver and brain biochemistry in studies using similar samples. Untreated NPC felines have previously showed an elevated hepatic alanine amino-transferase (ALT) and cholesterol, as well as decreased albumin in serum compared

to those of control felines [207]. Body weight has also been noted to be lowered in untreated NPC felines *versus* controls [207]. Liver biochemistry showed elevated unesterified cholesterol, sphingomyelin, bis(monoacylglycero)phosphate (BMP), glucosylceramide, lactosylceramide, globotriaosylceramide, free sphingosine and GM3 ganglioside in NPC felines over those of controls [207]. Biochemical brain changes have been previously observed in NPC felines against a control groups [207]. GM2 elevated substantially at 4 weeks of age from 9.5% of the total ganglioside content to 14% at 11 weeks [207]. However, elevations in GM3 were noted later in felines, and at a 4 week age point were substantially lower at 5.3% [207].

Interestingly, NPC felines treated pre-symptomatically were neurologically normal until the age of 76 weeks, at which point they were exhibiting mild ataxia [207]. As a result, the observations made here may provide early markers of this disease in felines. In the brains of these felines, unesterified cholesterol and GM2 gangliosides markedly reduced in concentration [207]. However, there was no improvement in serum albumin, and ALT and cholesterol concentrations were only slightly decreased when compared to untreated NPC cats [207]. Moreover, lipid biochemistry and liver histology were similar to the untreated felines [207]. This corroborates with the findings found in the feline urinary profiles in this study, in which no significant differences in the downstream products in feline metabolism were observed when comparing the profiles of early treated NPC felines to the untreated NPC ones.

However, felines treated with 1,000 mg SQ/120 mg IT were shown to be normal at 24 weeks, and only mild ataxia at 76 weeks of age [207]. This is consistent with the findings observed here in the urinary profiles. Unlike the felines administered with 120 mg beginning at 3 weeks, the levels of ALT and albumin in blood serum were consistent with normal felines [207]. Furthermore, serum cholesterol had significantly reduced when compared to untreated felines [207]. The liver showed significant improvements, such as decreased cholesterol, sphingomyelin, GM3 ganglioside, neutral glycolipids, and free sphingosine concentrations [207]. As significant differences are being observed in the plasma of these animals, this ultimately will also exert an impact on the feline urinary profiles. This further corroborates why significant differences are being observed in the feline urinary profiles that have been treated with combined 1,000 mg SQ /120 mg IT treatment, rather than 120 mg IT alone pre- or post-symptomatically.

Drug Detection in ^{1}H NMR

Despite some of the felines taking antibiotics, the parent drug and corresponding metabolites were not administered in doses which are detectable by ^{1}H NMR analysis (Appendix 8). Furthermore, HPβCD was not NMR-observable in the urine (Appendix 8), and this observation is consistent with other studies that demonstrate HPβCD is rapidly cleared [202], [235].

Implications of Experimental Design

Implications of this study which provide some limitations of the results acquired are that the NPC cohort is predominantly female, and also the overall small sample size in this investigation. More samples would be required in order to validate the findings that the 1,000 mg SQ / 120 mg IT treatment is the most effective one for NPC felines, and also perhaps a human NPC disease population. This study is limited by these two factors, and in order to validate these findings, a more balanced cohort of female to male ratio, and more samples could be used in order to substantiate the findings made here. The significant differences observed could be amplified in view of these age and sex factors.

Alternative proposed biomarkers for NPC disease

Bile acids (BAs), which have been previously recognised as significant biomarkers [13], were not identified in the urinary profiles (δ = 0.50-0.75 ppm). This is most likely attributable to the NMR sensitivity used; indeed, 400 MHz profiles were acquired in this case whereas in the study of Ruiz-Rodado *et al.* (2014), a 600 MHz NMR facility was employed [13]. These markers were found to be significant in view of their nature as downstream products of cholesterol. Hence, 600 MHz profiles of feline urine, in particular observing the felines with NPC to observe BAs. However, no characteristic BA resonances in the ^{1}H NMR profiles of NPC feline urine were obtained when employing an operating frequency of 600 MHz. (Appendix 11). Therefore, further metabolomics analysis using the 600 MHz facility were not performed and it maybe that BAs are not observed due to the early stage of the disease, as humans with NPC live much longer respectively. Another way to observe BAs using NMR is to complete an extraction using tetra-deutero methanol [218]. This could be observed as a further work for potential early biomarker detection in felines; however, these bile acid biomarkers have been established in human trials [13].

3-Aminoisobutyrate was not found to be a significant metabolite, which was suggested as a biomarker for human NPC patients using multivariate analysis [13]. However, it is challenging to compare the human studies with the studies of the felines, since the 13 untreated patients with NPC1 and 47 heterozygote carriers used in the trial were aged 2.7-30

and 25-62 years respectively, and therefore presumably the course of the disease will have been more manifested in these patients. Furthermore, comparison of these studies is also challenging in view of differences in metabolic pathways in the respective species.

4.9 Conclusions

^{1}H NMR metabolomics analysis of feline urine has been performed for the first time on treated and untreated NPC1 disease and corresponding control animals. Significant differences in profiles in older felines were observed when the disease has further progressed. However, in the felines that are younger, these differences were not statistically significant. This would suggest that ^{1}H NMR analysis of feline urine for early biomarker detection is unsuitable in felines. Nevertheless, treatment outcomes can be assessed in feline urine, with significant differences observed in NPC comparatively to treated 120 mg HPβCD IT and 1,000 mg/kg SQ felines. Metabolites that were observed to be significant indicate disturbances in the gut microbiota and sterol/cholesterol metabolism pathways which is in-line with the NPC disease pathology.

Chapter 5: ^{1}H NMR Metabolomic Profiling of GM1 Gangliosidosis (Type II) Patients: Potential Diagnostic Applications from Urinary and Blood Plasma Screens

5.0 Abstract

Background: GM1 Type II is a rare, autosomal, neurodegenerative disorder that arises from a mutation in the GLB1 gene. GM1 Type II has a range of symptoms, including but not limited to neurological defects, hepatosplenomegaly and skeletal abnormalities. Although some metabolomic investigations have been previously performed, a full global screening of metabolites in urine, plasma and cerebrospinal fluid (CSF) from these patients, and its biomarker potential, is yet to be explored.

Methodology: GM1 Type II plasma, urine and CSF samples were collected at the National Institute of Health, and corresponding healthy control plasma and urine were collected from De Montfort University. ^{1}H nuclear magnetic resonance (NMR) profiles were acquired using a 700 MHz NMR facility employing *noesypresat* and *cpmg* pulse sequences for these biofluids. The ^{1}H NMR profiles were assigned and then processed into excel data matrices. Univariate (UV) and multivariate (MV) statistical analysis strategies were employed on datasets acquired. The identification of GM1 gangliosides using liquid chromatography-mass spectrometry (LC-MS) was also evaluated in the biofluids collected.

Results and Discussions: GM2 gangliosides was not detectable using the LC-MS strategy employed on all GM1 Type II biofluids. Significant statistical differences in plasma and urinary profiles using ^{1}H NMR profiles both in MV and UV strategies. Statistically significant metabolites include urinary upregulated citrate, N-acetyls, creatinine, creatine and lactate amongst other metabolites. Additionally, in plasma downregulated low-density lipoprotein (LDL) cholesterol and upregulated creatinine and lactate amongst other metabolic disturbances was observed. However, interpretation of these results is treated with caution, due to the different ages and fasting status in the two respective cohorts. The impact of xenobiotics, such as drugs and supplements, that the GM1 Type II cohort were exposed to as part of this investigation is also considered. This study reveals the complexity of analysing lysosomal storage disorders using a human model using a global untargeted approach.

Conclusions: In this pilot study, it is challenging to validate GM1 Type II biomarkers, with so few samples available and due to the critical status of the patients to ensure equivalent conditions. Indeed, in combination with these variables and no treatment to validate findings

this presents a challenge in metabolomic investigations. Future validation studies could be performed, or the use of an animal model to create perfect conditions, in order to confirm findings. The potential for TAGs and LDL amongst other metabolites as unspecific markers in GM1 Type II plasma and n-acetyl groups, lactate, creatinine amongst other metabolites to be unspecific markers in GM1 Type II urine. This is consistent with other lysosomal storage disorders and the literature.

Keywords: Nuclear Magnetic Resonance; Metabolomics; Lysosomal Storage Diseases; Gangliosidosis GM1 Type II; Blood Plasma; Urine; Biomarkers

5.1 Introduction

GM1 Type II (*#OMIM 230600*) is the juvenile form of GM1 disease, with onset typically occurring between in the first few months to years of life, and overall life expectancy is early adulthood [236]. Typical symptoms of this disorder include ataxia, seizures, dementia, difficulty with speech, skeletal abnormalities and hepatosplenomegaly [236]. The disease arises from a mutation in the GLB1 gene. At present, there is no established cure or treatment for GM1 Type II disease, although medication is frequently prescribed symptomatically, i.e. anticonvulsants to control associated seizures.

As highlighted in the literature review (Chapter 1), to date few metabolomics investigations have been performed. In principle, such investigations have the potential to provide candidate biomarkers for this debilitating condition, together with a further understanding of the abnormal biochemistry mediated by GM1 Type II disease. Indeed, it has previously been demonstrated that N-acetylated biantennary octasaccharide, N-acetylated triantennary decasaccharide/tetraantennary dodecasaccharides are present in GM1 Type I (*OMIM 230500*) urinary profiles using thin- layer chromatography (TLC) coupled with ^{1}H NMR analysis [237]. GM1 Type I represents the most severe type form of gangliosidosis, with patient life expectancies within the childhood ranges, so it might be expected that this form has the highest possible level metabolic disturbances. GM1 Type III is the adult form of the condition. However, this investigation was targeted on these N-acetyl compounds alone, and not the global metabolome, in which a combination of markers could be attributable to GM1 Type I. Metabolomics investigations have also observed the quantification of gangliosides, potentially key biomarkers in GM1 and GM2 gangliosidoses, *in vitro* but not *in vivo* [48].

In this investigation, conventional ^{1}H NMR- and LC-MS-based metabolomics approaches were employed in order to identify new, potentially valuable biomarkers for GM1 Type II gangliosidosis, particularly those which correlate with their severity and progression. In view of the age differences in the biofluid samples collected (blood plasma, CSF and urine) the abilities of these techniques to prognostically stratify such patients was also explored. Furthermore, these strategies were supplemented by the application of newly-developed metabolic pathway and quantitative metabolite set enrichment analyses; these software-based approaches serve to provide much valuable information regarding biomolecular connectivities, and hence the biological/metabolic pathways involved in the pathogenesis of GM1 diseases. This approach also permitted the recognition of metabolic 'signatures' that may enhance our understanding of the pathogenic mechanisms of GM1 at the molecular, biochemical and cellular levels. Therefore, the overall approach employed combined both targeted and untargeted metabolomics strategies.

5.2 Specific Aims

- Discovery of age-related changes arising from perturbed metabolic pathways which reflect pathogenic mechanisms associated with GM1 Type II gangliosidosis.
- To seek and identify new biomarkers of GM1 Type II disease activities and progression in blood plasma, CSF and/or urine.
- Employment of these multidimensional metabolic datasets in order to enhance our understanding of GM1 type II disease processes at the molecular, biochemical and cellular levels

5.3 Approval for Research

The investigation was carried out abiding by the rules of the Declaration of Helsinki 1975 (revised in 2013). All subjects gave their informed consent for inclusion before they participated in the study. De Montfort University, Leicester provided ethical approval for the protocol of collection of plasma from control participants (reference no. 1936). The GM1 Type II patients were consented under protocol 02-HG-0107 "Neurodegeneration in Glycosphingolipid Storage Disorders" under the National Human Genome Research Institute IRB.

5.3 Sample Collection of Human Biofluids

10 matched blood plasma, urine and CSF specimens collected from groups of GM1 Type II patients were collected by Prof. Tifft's team based at the National Institute of Health (NIH), Bethesda, Maryland, 20892, USA, and were transported on dry ice to Leicester School of Pharmacy, De Montfort University (DMU), Leicester, UK, and then immediately frozen at-80°C on arrival at the laboratory. All samples were collected from different individuals, with the exception of one sample, which was derived from a male at two different time points (i.e., when this patient was aged 18 and 20 years).

Blood plasma samples were not discoloured, and therefore it was presumed that no blood contamination had occurred; therefore, all samples were included in the ^{1}H NMR and LC-MS analysis regimens. A range of samples were collected from different age ranges and individuals in order to gain a longitudinal insight into the molecular mechanisms featured in the disease process. Blood plasma was collected into heparinized tubes in order to avoid NMR analytical complications arising from the use of collection tubes containing an EDTA anticoagulant discussed in Chapter 3.

Class	GM1 Type II	Controls	Total
Mean Age	11.4	22.4	20.5
Range (Years)	(3-20)	(18-54)	(3-54)
Males	4	10	14
Females	6	27	33
Total	10	37	47

Table 5.0. Details and breakdown of GM1 Type II blood plasma and corresponding healthy control samples, i.e. age and gender for both groupings.

Class	GM1 Type II	Controls	Total
Mean Age	11.4	40.4	37
Range (Years)	(3-20)	(18-68)	(3-68)
Males	4	25	29
Females	6	51	57
Total	10	76	86

Table 5.1. Details and breakdown of GM1 Type II urine and corresponding healthy control samples, i.e. age and gender for both groupings.

CSF was only collected from the 10 GM1 Type II patients and not control participants. GM1 Type II patients were on a combination of drugs, some of which have ^{1}H NMR signals which

overlap with metabolite resonances in the ^{1}H NMR profiles acquired. Since all participants were on a range of medications, as many drug metabolite signals as possible were determined prior to performing multivariate statistical analysis, in order to address their potential impact on the overall metabolomes.

5.4 Preparation of Human Biofluids for ^{1}H NMR Analysis

Biofluid samples were thawed at room temperature and then immediately prepared for ^{1}H NMR analysis. Preparation for this set of samples involved centrifuging 500 µl volumes of sample (urine or plasma) and removing 350 µl supernatants for analysis. Degradation of CSF and plasma has been observed when left at ambient temperature for periods of more than 2.5 hours [238], so these samples were prepared and analysed immediately. A 50 µl aliquot of phosphate buffer (Acros Organics) at pH 7.00 was added to the supernatant, including 0.05% (w/v) sodium azide (50 µl) (Sigma-Aldrich), and then 10% (v/v) D_2O (45 µl) (Sigma Aldrich), also containing 0.05% (w/v) 3-(trimethylsilyl)propionic-2,2,3,3-d4 acid, sodium salt (TSP) (Sigma Aldrich) was then added to the solution. For the plasma samples, the internal quantitative standard TSP was not added since its resonance can broaden substantially from binding to plasma proteins such as albumin. These mixtures were then thoroughly rotamixed and added to newly-purchased NMR tubes (Norell) ready for analysis.

5.5 ^{1}H NMR Analysis

Samples were analysed at the University of Oxford using the Bruker AVII 700 MHz NMR spectrometer equipped with a ^{1}H TCI cryoprobe (Department of Chemistry) at an operating frequency of 699.989 MHz and a probe temperature of 298 K. Spectral acquisition involved the collection of 65,536 data points using 32 scans across a spectral width of 15.99 ppm. For the urine, plasma and CSF samples, NOSEY Presaturation (noesygppr1d) was used as previously described (Chapter 2), targeting the intense water signal located at $\delta = 4.70$ ppm. For the plasma and CSF samples, the CPMG pulse sequence was employed in order to suppress the broad protein signal envelope which obscures the visibility, and hence quantification, of many low-molecular-mass metabolite resonances. Samples were loaded onto an automated sample belt using a sample changer.

5.6 NMR Data Preprocessing

^{1}H NMR data were preprocessed in the same manner as described in Chapters 3 and 4 for both plasma and urine samples. The removal of bucket regions attributable to drug ^{1}H NMR signals, and those of their metabolites, is described below in section.

5.7 Liquid Chromatography-Mass Spectrometry

Using an Agilent 6120 quadrupole LC-MS system, a method was developed to identify GM2 ganglioside (1383.4 m/z) in negative ion mode using signal ion monitoring (SIM) parameters. High-performance liquid chromatographic (HPLC) analysis was performed on an Agilent 1260 infinity system, using a Supelcosil LC-18-DB, 3 μm, 3.3 cm x 4.6 mm column. 20 μl of the authentic GM2 standard was injected at a flow-rate of 1.00 ml/min. employing the following gradient elution profile: 0.00-2.50 min. 100% (C), 2.51 min. 10% (A) and 90% (B), 5 min. 100% (B) [where A is water/methanol 98:2% (v/v)]; B: 100% (v/v) methanol; and C 92:8% (v/v) water/methanol]. A signal observed at a retention time of 0.54 min. represented GM2 ganglioside (1383.4 m/z).

5.8 ^{1}H NMR Profiles of Plasma and Urine from GM1 Type II Patients

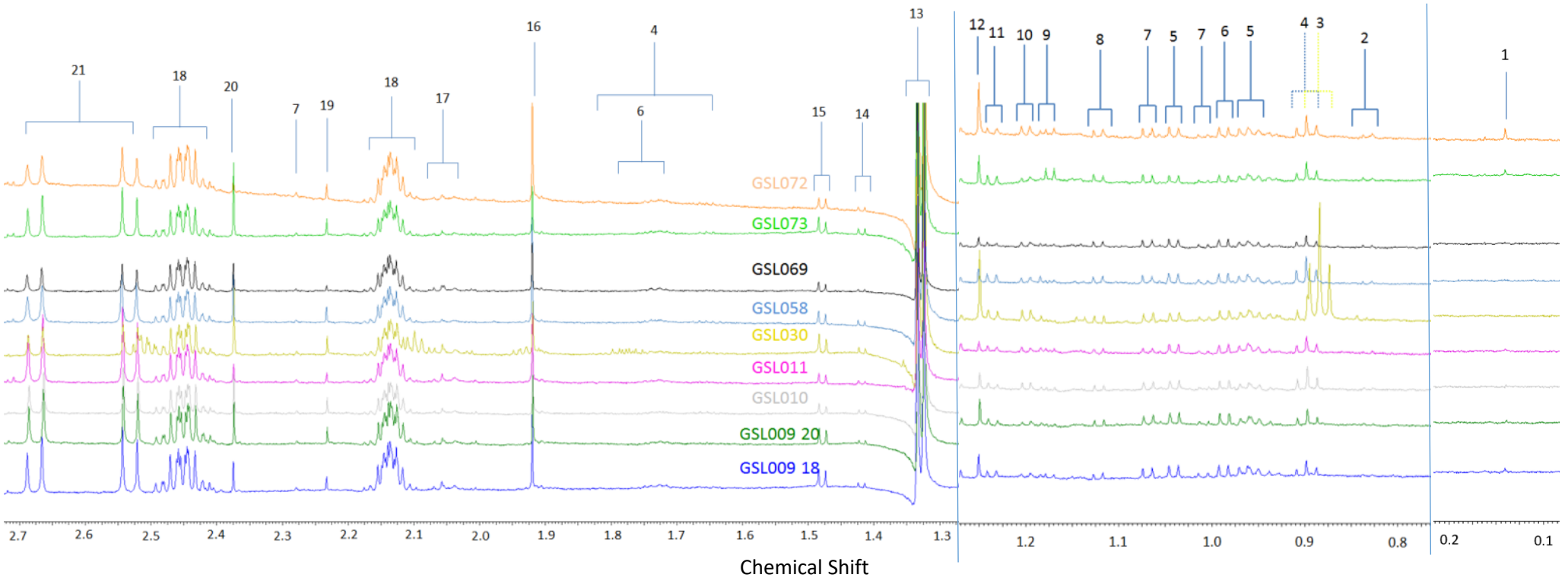


Figure 5.0. (a) Partial Profiles of CPMG ^{1}H NMR spectra of GM1 Type II CSF in the 0.10 ppm-2.70 ppm range. Assignments: [1] Silicone grease-C<u>H</u>$_3$; [2] 2-Hydroxy-3-methylbutyrate-C<u>H</u>$_3$; [3] Levetiracetam-C<u>H</u>$_3$; [4] 2-Hydroxybutyrate-C<u>H</u>$_3$; [5] Isoleucine-C<u>H</u>$_3$; [6] Leucine-C<u>H</u>$_3$; [7] Valerate-C<u>H</u>$_3$; [8] α-ketoisovalerate-C<u>H</u>$_3$; [9] Methylsuccinate-C<u>H</u>$_3$; [10] Propylene glycol-C<u>H</u>$_3$; [11] 3-D-Hydroxybutyrate-C<u>H</u>$_3$; [12] 2,2-Dimethylsuccinate-C<u>H</u>$_3$; [13] Lactate-C<u>H</u>$_3$; [14] ^{13}C satellite; [15] Alanine-C<u>H</u>$_3$; [16] Acetate-C<u>H</u>$_3$; [17] N-Acetyl-function-C<u>H</u>$_3$ of acute-phase glycoprotein molecularly-mobile carbohydrate side-chain N-acetylsugars; [18] Glutamine-β- and γ-C<u>H</u>$_2$ functions (2.12 and 2.45 ppm respectively); [19] Acetone-C<u>H</u>$_3$; [20] Pyruvate-C<u>H</u>$_3$; [21] Citrate-C<u>H</u>$_{2A}$/C<u>H</u>$_{2B}$.

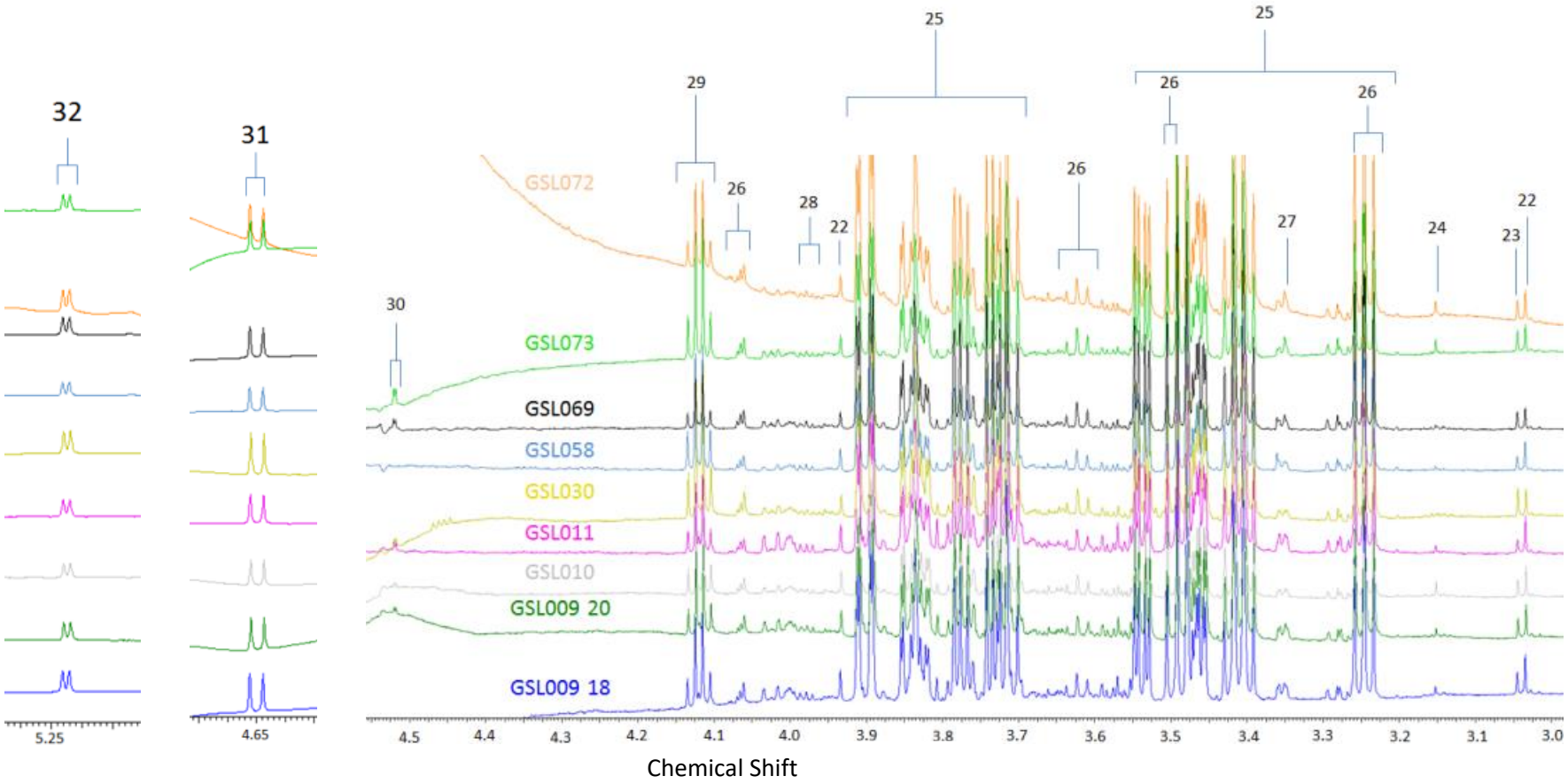


Figure 5.0. (b) Partial CPMG ^{1}H NMR spectra of GM1 Type II CSF in the 2.70 ppm – 5.20 ppm range. Assignments: [22] Creatine-C<u>H</u>$_3$; [23] Creatinine-C<u>H</u>$_3$; [24] Dimethyl Sulfone-C<u>H</u>$_3$; [25] D-Glucose bulk chain-C2-C6-C<u>H</u>; [26] myo-Inositol-<u>H</u>/C2/C6,C4/C3,C1/C5; [27] Methanol-C<u>H</u>$_3$; [28] 2-Hydroxybutyrate-C<u>H</u>; [29] Lactate-C<u>H</u>; [30] tentative Ascorbate-C<u>H</u>; [31] β-Glucose-C1-C<u>H</u>; [32] α-Glucose-C1-C<u>H</u>.

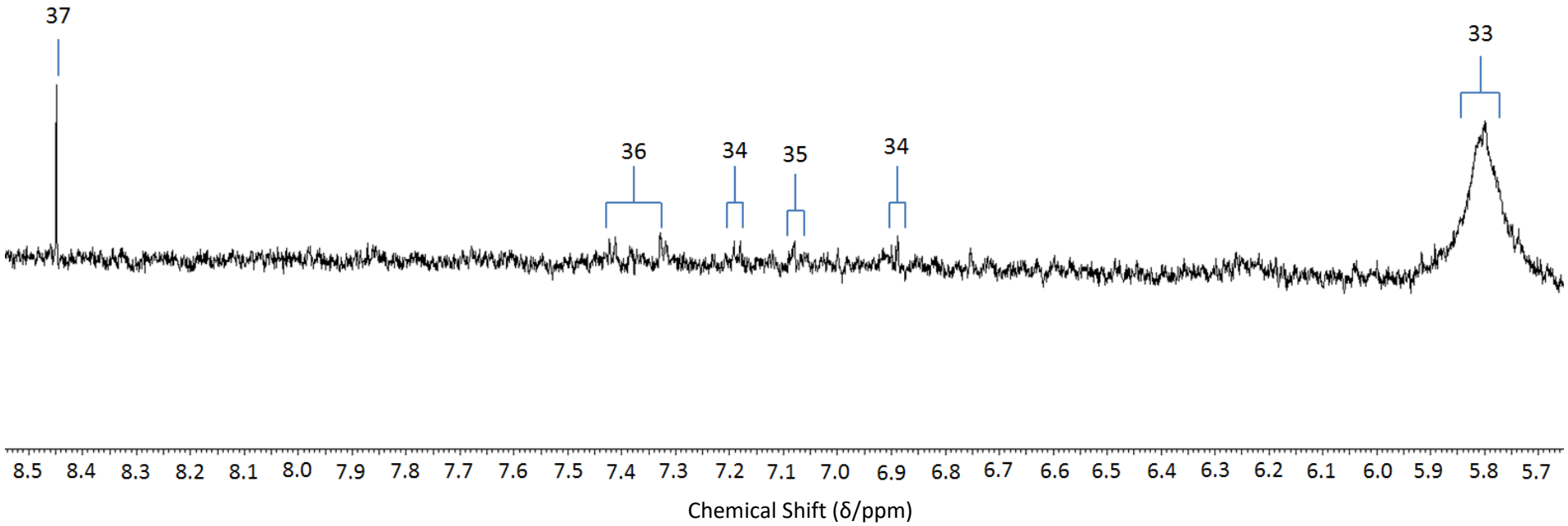


Figure 5.0. (c) Partial CPMG ^{1}H NMR spectrum of GM1 Type II CSF (sample code GSL011) in the 5.65 – 8.55 ppm range. Assignments: [33] Urea-CO-N$\underline{H}_2$; [34] L-Tyrosine-C3/C5-C$\underline{H}$ (trace levels detectable); [35] unassigned; [36] L-Phenylalanine-C4/C5,C3/C6,C2-$\underline{H}$ (trace levels detectable); [37] Formate-C$\underline{H}$.

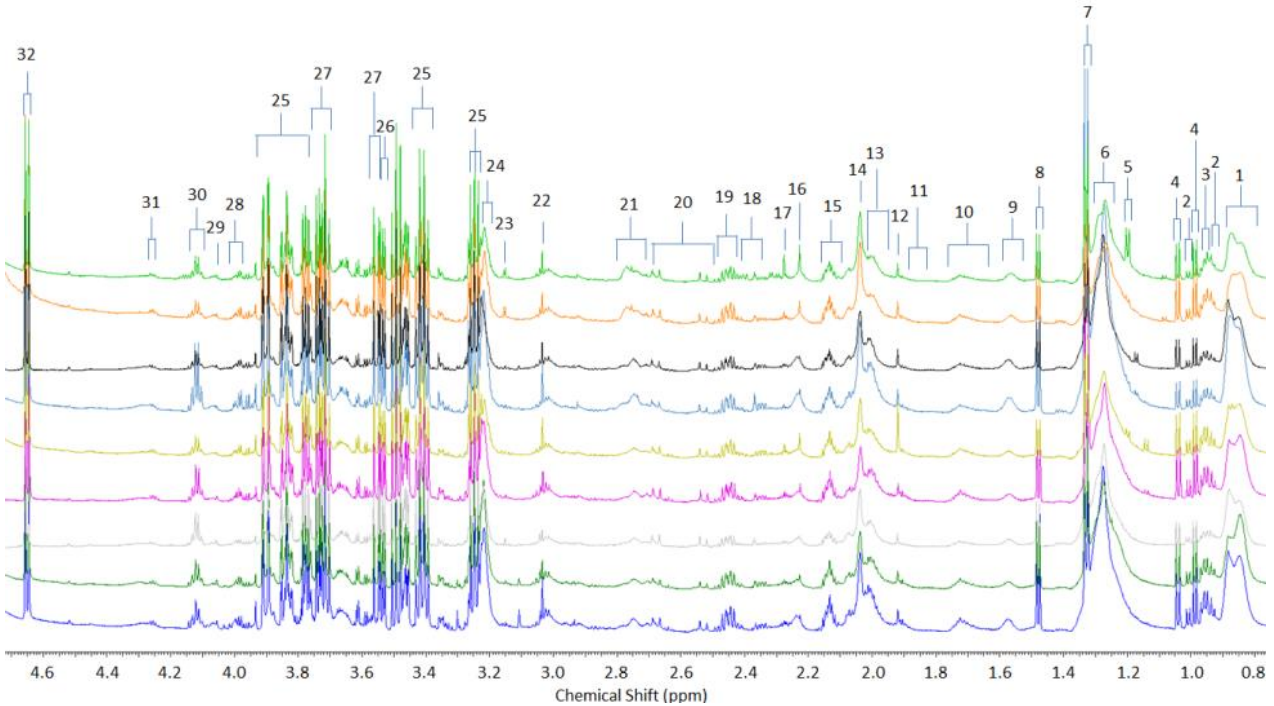


Figure 5.1. (a) Partial CPMG ^{1}H NMR spectrum of typical blood plasma sample collected from GM1 Type II patients (0.80 -4.70 ppm range). Assignments: [1] Very-low-density-lipoprotein (vLDL) /low-density-lipoprotein (LDL)-triacylglycerol (TAG)-C$\underline{H}_3$s; [2] Isoleucine-C$\underline{H}_3$; [3] Leucine-C$\underline{H}_3$; [4] Valine-C$\underline{H}_3$; [5] 3-D-Hydroxybutyrate-C$\underline{H}_3$; [6] vLDL/LDL-bulk chain-(C$\underline{H}_2$)$_n$; [7] Lactate-C$\underline{H}_3$; [8] Alanine-C$\underline{H}_3$; [9] TAG–C$\underline{H}_2$-CH$_2$CO; [10] Lysineγ-C$\underline{H}_2$/Arginine-C$\underline{H}_2$; [11] Arginine-γ-C$\underline{H}_2$; [12] Acetate-C$\underline{H}_3$; [13] Lipoprotein-TAG-C$\underline{H}_2$-CH=CH-; [14] N-Acetyl-function-C$\underline{H}_3$; [15] Glutamine-β-C$\underline{H}_2$;[16] Acetone-C$\underline{H}_3$; [17] Acetoacetate-C$\underline{H}_3$; [18] Glutamate–γ-C$\underline{H}_2$; [19] Glutamate–γ-C$\underline{H}_2$; [20] Citrate-C$\underline{H}_{2A}$/-C$\underline{H}_{2B}$; [21] TAG-CH=CH-C$\underline{H}_2$-CH=CH; [22] Creatine/Creatinine-C$\underline{H}_3$; [23] unassigned; [24] Choline headgroup of high density lipoprotein (HDL) phospholipids (-N$^+$(C$\underline{H}_3$)$_3$); [25] Glucose-C$\underline{H}$; [26] Glycine-C$\underline{H}_2$; [27] Glucose-Bulk chain-C2-C6-$\underline{H}$; [28] Tyrosine-C$\underline{H}$/Phenylalanine-C$\underline{H}$/L-Histidine-α-C$\underline{H}$; [29] Creatine-CH$_2$; [30] Lactate-C$\underline{H}$; [31] L-Threonine-C$\underline{H}$/3-D-Hydroxybutyrate-C$\underline{H}_3$; [32] β-Glucose-C1-C$\underline{H}$.

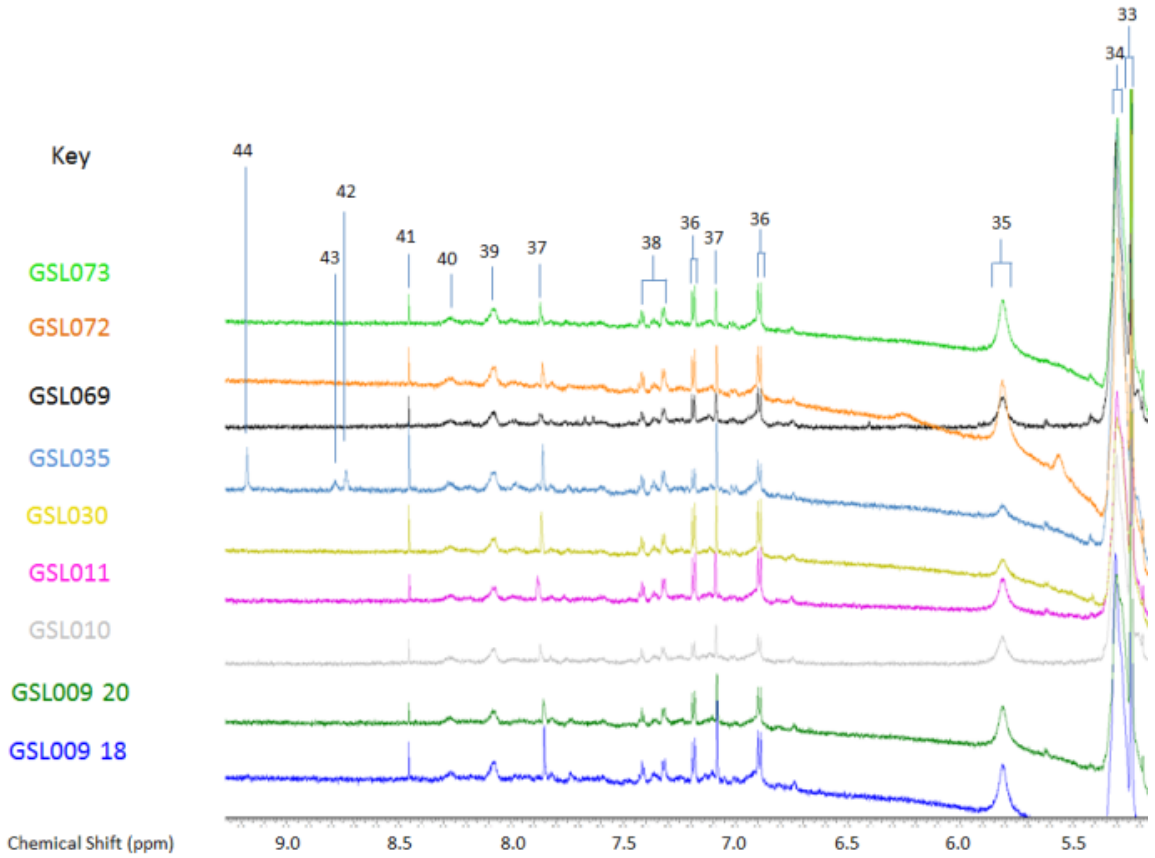
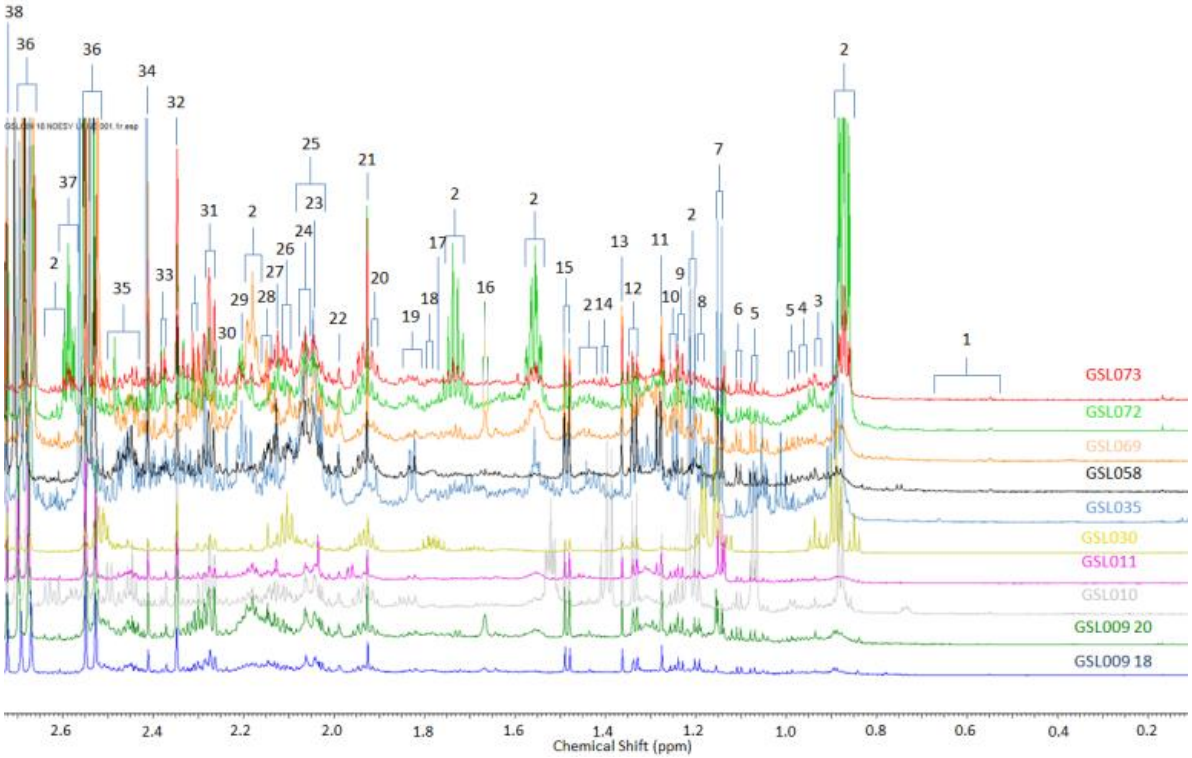


Figure 5.1. (b) Partial ^{1}H NMR spectra of blood plasma samples collected from GM1 Type II patients. Typical spectra are shown (5.00-9.50 ppm and 14.10-14.50 ppm regions) Assignments: [33] α-Glucose-C1-CH; [34] TAG-C<u>H</u>=C<u>H</u>-; [35] Urea-CO-N<u>H</u>$_2$; [36] L-Tyrosine aromatic-C3/C5-C<u>H</u>; [37] L-Histidine-C2-C<u>H</u>; [38] L-Phenylalanine aromatic-C4/C5,C3/C6,C2-<u>H</u>; [39] Protein aromatic amino acid-C<u>H</u> (tentative assignment); [40] Protein aromatic amino acid-CH (tentative assignment); [41] Formate-C<u>H</u>; [42] Trigonelline-Cδ/Cζ-C<u>H</u>; [43] Trigonelline-Cβ-C<u>H</u>; [44] unassigned.


GSL073
GSL072
GSL069
GSL058
GSL035
GSL030
GSL011
GSL010
GSL009 20
GSL009 18
Chemical Shift (ppm)

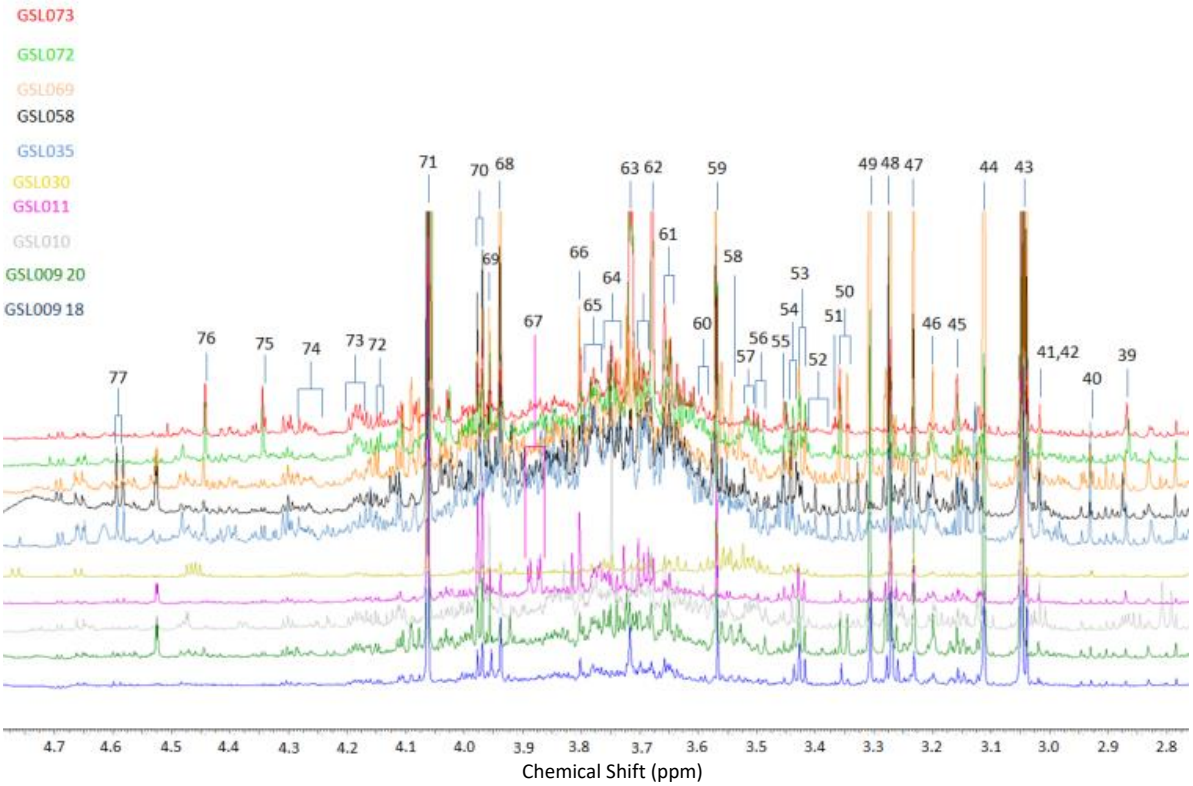
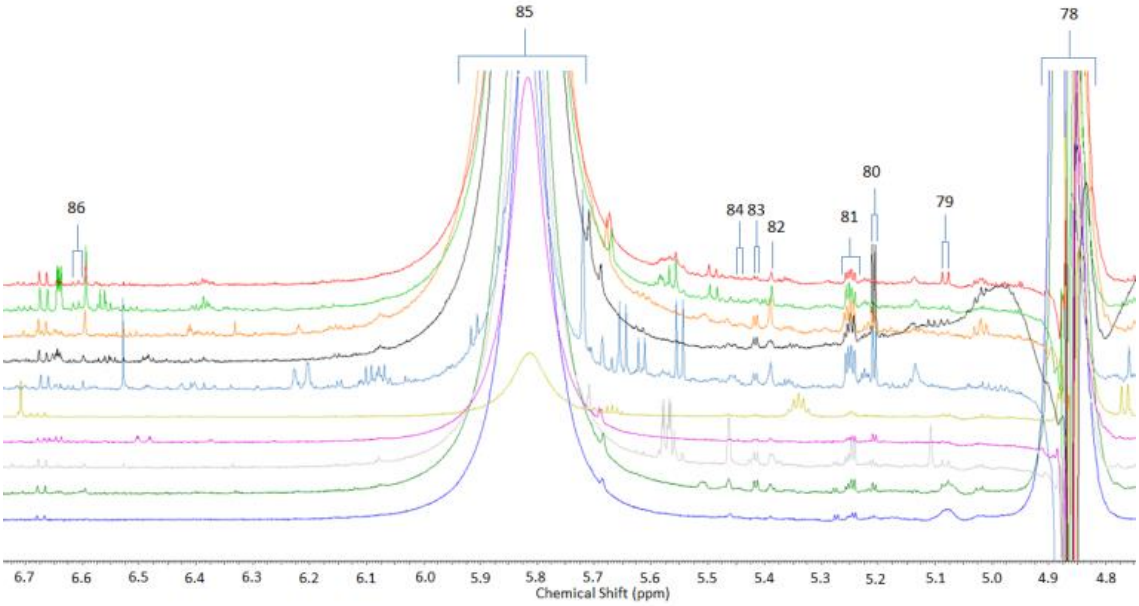

GSL073
GSL072
GSL069
GSL058
GSL035
GSL030
GSL011
GSL010
GSL009 20
GSL009 18
Chemical Shift (ppm)

Figure 5.2. (a) Partial NOESY single-pulse ^{1}H NMR spectrum of GM1 Type II patient urine (0.85-4.70 ppm range). Typical spectra are shown. Assignments: [1] Bile acid-C$\underline{H}_3$s; [2] Tentative lamotrigine metabolite and miglustat parent drug resonances-C$\underline{H}_3$; [3] Isoleucine-C$\underline{H}_3$; [4] Leucine-C$\underline{H}_3$; [5] Valine-C$\underline{H}_3$s; [6] Methyl Succinate-C$\underline{H}_3$; [7] 2-Oxoisovalerate-C$\underline{H}_3$; [8] Aminoisobutyrate-C$\underline{H}_3$; [9] 3-D-Hydroxybutyrate-C$\underline{H}_3$; [10] L-Fucose-C$\underline{H}_3$; [11] 3-Hydroxyisovalerate-C$\underline{H}_3$; [12] Lactate-C$\underline{H}_3$/Threonine-C$\underline{H}_3$; [13] 2-Hydroxybutyrate-C$\underline{H}_3$; [14] Lysine-Cγ-C$\underline{H}_2$; [15] Alanine-C$\underline{H}_3$; [16] Unassigned; [17] Lysine-Cβ-C$\underline{H}_2$; [18] Thymine-ε-C$\underline{H}_3$; [19] Aminoadipate-C$\underline{H}_2$; [20] Lysine-γ-C$\underline{H}_2$; [21] Acetate-C$\underline{H}_3$; [22] 2-Hydroxyglutarate-C$\underline{H}_2$; [23-25] N-Acetyl-C$\underline{H}_3$ functions; [26] Glutamate-β-C$\underline{H}_2$; [27] Acetoacetate-C$\underline{H}_3$; [28] Glutamine-γ-C$\underline{H}_2$; [29] Acetoin-α-C$\underline{H}_3$; [30] Acetoacetate-C$\underline{H}_3$; [31] 2-Hydroxyglutarate-δ-C$\underline{H}_2$; [32] Pyruvate-C$\underline{H}_3$; [33] Glutamate-γ-C$\underline{H}_2$; [34] Succinate-C$\underline{H}_2$'s; [35] Glutamine-δ-C$\underline{H}_2$; [36] Citrate-C$\underline{H}_{2A}$/C$\underline{H}_{2B}$; [37] tentative Lamotrigine Resonance-C$\underline{H}$; [38] Dimethylamine-C$\underline{H}_3$; [39] Trimethylamine-C$\underline{H}_3$; [40] Dimethylglycine-C$\underline{H}_3$; [41] L-Lysine-ε-C$\underline{H}_2$; [42] L-Tyrosine-β-C$\underline{H}$; [43] Creatinine-C$\underline{H}_3$/Creatine-C$\underline{H}_3$; [44] Malonate-C$\underline{H}_2$; [45] L-Histidine-β-C$\underline{H}_2$; [46] D-Glucose-C$\underline{H}$ [47] TMAO-C$\underline{H}_3$; [48] Taurine-C$\underline{H}_2$-NH_3^+; [49] Trimethylamine-C$\underline{H}_3$; [50] Glucuronate-γ-C$\underline{H}$; [51] Methanol-C$\underline{H}_3$; [52] D-Glucose-C$\underline{H}$; [53] Cystine-γ/δ-C$\underline{H}_2$ [54] cis-Aconitate-C$\underline{H}_3$; [55] Taurine-C$\underline{H}_2$-SO_3^-; [56] 3-Hydroxyphenylacetate-C$\underline{H}_2$; [57] Phenylacetate-C$\underline{H}_2$; [58] unassigned; [59] Glycine-C$\underline{H}_2$; [60] Phenylacetylglycine-γ-C$\underline{H}_2$;/Threonine-β-C$\underline{H}$; [61] Mannitol-C$\underline{H}_2$; [62] Phenylacetylglutamine-β-C$\underline{H}_2$; [63] Unassigned; [64] Lysine-ε-C$\underline{H}$; [65] Ethanolamine-C$\underline{H}_2$; [66] Mannitol-C$\underline{H}_2$OH; [67] Mannitol-C$\underline{H}_2$; [68] Creatine-C$\underline{H}_2$; [69] Glycolate-C$\underline{H}_2$; [70] Hippurate-C$\underline{H}_2$; [71] Creatinine-C$\underline{H}_2$; [72] Pyroglutamate-β-CH; [73] Lactate-C$\underline{H}$; [74] Threonine-β-C$\underline{H}$; [75] Tartrate-C$\underline{H}$OH; [76] Trigonelline-C$\underline{H}_3$; [77] β-Glucose-C1$\underline{H}$.


85
78
86
84 83
82
81
80
79
6.7 6.6 6.5 6.4 6.3 6.2 6.1 6.0 5.9 5.8 5.7 5.6 5.5 5.4 5.3 5.2 5.1 5.0 4.9 4.8
Chemical Shift (ppm)

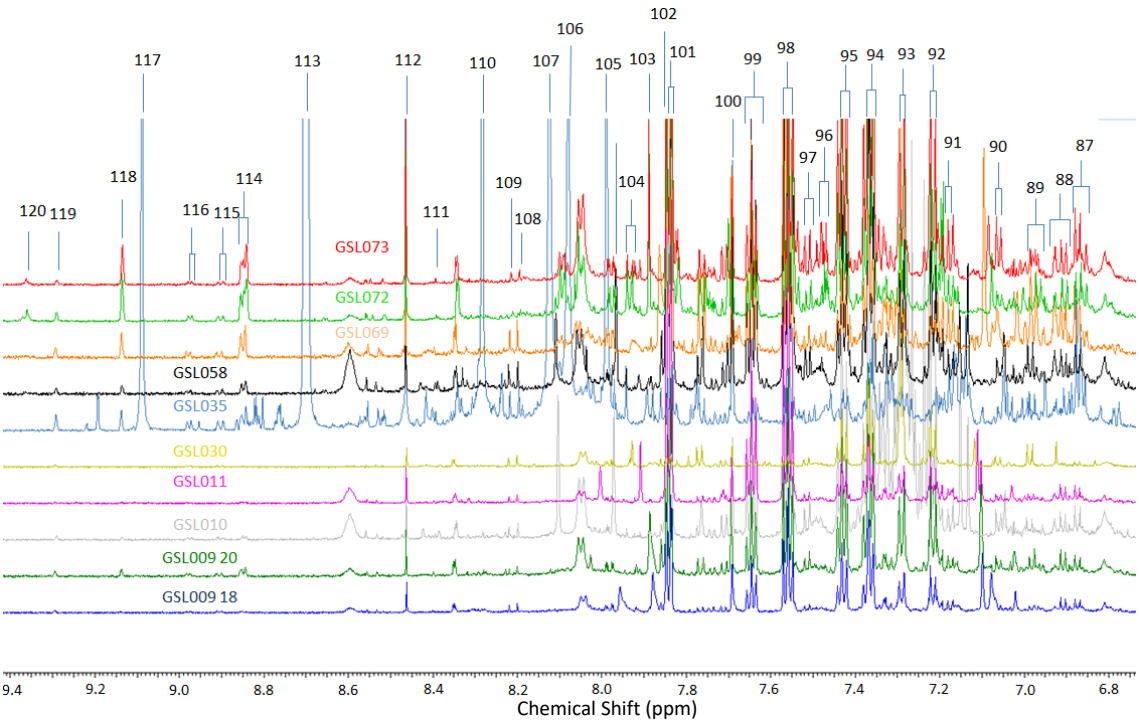

102
106
117
113
112
110
107
105
103
101
99
98
95 94 93 92
100
96
97
91
90
87
118
114
109
104
89
88
120 119
116 115
111
108
GSL073
GSL072
GSL069
GSL058
GSL035
GSL030
GSL011
GSL010
GSL009 20
GSL009 18
9.4 9.2 9.0 8.8 8.6 8.4 8.2 8.0 7.8 7.6 7.4 7.2 7.0 6.8
Chemical Shift (ppm)

Figure 5.2. (b) Partial NOESY single-pulse ^{1}H NMR spectrum of GM1 Type II patient urine (4.70-9.30 ppm range). Typical spectra are shown. Assignments: [78] Residual water-OH; [79] Unassigned; [80] α-Glucose-CH; [81-82] Unassigned; [83] Sucrose-α-CH; [84] Unassigned; [85] Urea-CO-NH_2; [86] 2-Furoylglycine-β-CH; [87] Tyrosine-γ/ε-CH; [88] Unassigned [89] Phenol-β/δ/ζ-CH; [90] Phenol-γ/ε-CH; [91] Tyrosine-ζ/η-CH; [92] Indoxyl-sulphate-η/θ/β-CH; [93] Indoxyl-sulphate-ζ/ι-CH; [94] Indoxyl sulphate-β-CH and Phenylalanine-β/ζ-CH; [95] Phenylalanine-γ/δ/ε-CH; [96] Indoxyl-sulphate-β-CH; [97] Indoxylsulphate-ζ-CH and Benzoate-γ/ε-CH; [98] Hippurate-γ/ε-CH; [99] Hippurate-δ-CH; [100] 1-Methylhistidine-δ-CH; [101] Hippurate-β/ζ-CH; [102] Histidine-β-CH; [103] Histidine-CH; [104] 3-Methylhistidine-β-CH; [105-107] Tentative (tolperisone metabolite) drug metabolite-CH; [108] Hypoxanthine-β/η-CH; [109] 1-Methylnicotinamide-ε-CH; [110] Tentative (tolperisone metabolite) drug metabolite-CH; [111] Quinolinate-CH; [112] Formate-CH; [113] Tentative (tolperisone metabolite) drug metabolite-CH; [114] Trigonelline-δ/ζ-CH; [115] 1-Methylnicotinamide-ε-CH; [116] 1-Methylnicotinamide-ε-CH; [117] Tentative (tolperisone metabolite) drug metabolite-CH; [118] Trigonelline-β-CH; [119] N-Methylnicotinamide-β-CH; [120] Nicotinamide riboside- β -CH.

700 MHz single-pulse ^{1}H NMR spectra of these biofluids contained many prominent, sharp signals assignable to wide range of low molecular-mass biomolecules. Indeed, more than 80, 50 and 30 metabolites were detectable in the spectra acquired on urine, blood plasma and CSF respectively, and these included a range of biomolecules such as short-chain organic acid anions (e.g., acetate, formate, fumarate, lactate, pyruvate, succinate, glycolate and 3-D-hydroxybutrate); amino acids, including leucine, isoleucine and valine (collectively known as branched-chain amino acids, abbreviated BCAAs), glycine, alanine, glutamate, glutamine, lysine, proline, taurine, histidine, phenylalanine and tyrosine, etc., together with N-acetylamino acids; carbohydrates, most especially GM1 disease-relevant N-acetylsugars; and also nicotinate and nicotinamide pathway metabolites, amongst others (Figures 5.0-5.2). As expected, the ^{1}H NMR profiles of blood plasma contained relatively broad resonances arising from a series of lipoprotein-associated triacylglycerol's (TAG's), with ^{1}H NMR-distinguishable very low-, low- and high-density-lipoproteins, the acetamido (-$NHCOCH_3$) functions of N-acetylneuraminate and N-acetylglucosamine residues present in the molecularly-mobile carbohydrate side-chains of selected 'acute-phase' glycoproteins, in addition to those arising from aromatic amino acid protein residues.

In addition to the above resonances, those attributable to drugs and their corresponding metabolites were also assigned in the urine and CSF of these individuals. On closer inspection of the spectra acquired, such drug metabolites were assignable (Figure 5.3). These signals were removed prior to multivariate analysis (MVA). Full elucidation of the molecular structures of drug metabolites was not possible without the application of 2D NMR techniques.

Figure 5.3 shows a range of urinary xenobiotic metabolite resonances observed, for example those clearly attributable to valproate and its glucuronide metabolite; the full list of signals for these agents is already established in the literature [37]. Since only one participant is receiving each drug respectively, some metabolite signals attributable to drugs were retained in the statistical analysis model and it was firstly considered if these participant samples represented outliers in the preliminary, unsupervised forms of analysis first, i.e. PCA. However, the majority of the urinary bucket regions shown in Figure 5.3 were removed from the different forms of statistical analysis conducted, and these included the δ = 0.85 – 0.95, 5.54 – 5.56 ppm, 1.03 – 1.05 ppm, 1.37-1.39 ppm, 1.50 – 1.65 ppm, and 3.55 – 3.56 ppm regions, in addition to the xenobiotic signals highlighted in Figure 5.2, i.e. those at δ = 9.1, 8.7, 8.3, 7.9 and 8.05 – 8.15 ppm were all removed therefrom.

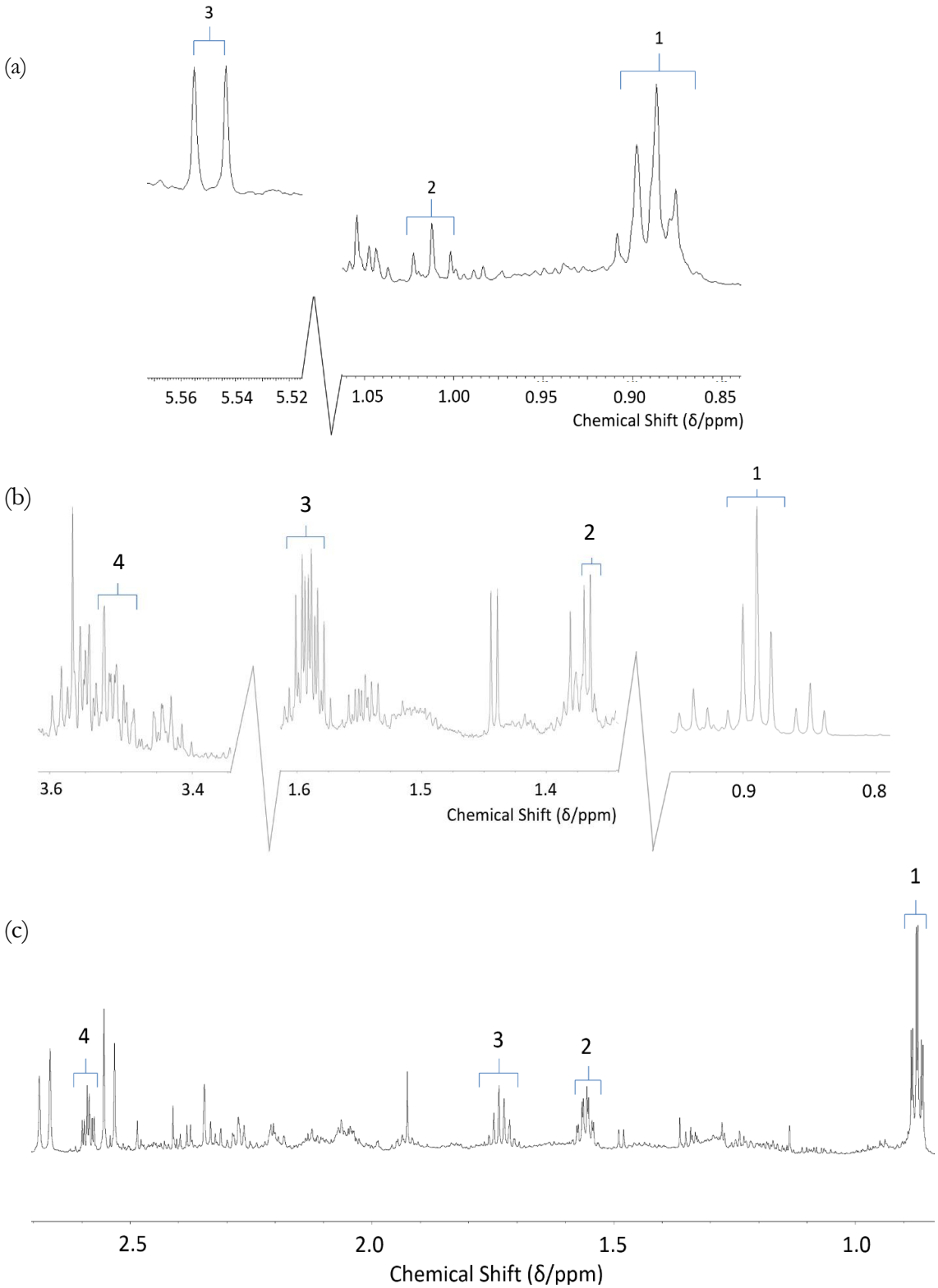


Figure 5.3. (a) Examples of resonances in the ^{1}H NMR profiles of human urine collected from GM1 type II patients of valproate metabolite 1-O-valproyl-β-glucuronide, including methyl protons [1] (*t*); [2] and anomeric proton (*d*); [3] References showing previous assignments of these resonances [37], [239]. (b) Example of resonances in the ^{1}H NMR profiles of human urine from GM1 Type II patients attributable to miglustat, including methyl protons [1]; β-CH_2 [2]; γ-CH_2 [3] and miglustat piperidine ring-γ/α-CH [4]. (c) [1]-[4] are resonances that arise from from drug administrations. Patient was receiving lamotrigine but these resonances do not correspond to this drug or its metabolites.

5.9 Statistical Analysis of Plasma Dataset

Multivariate Statistical Analysis

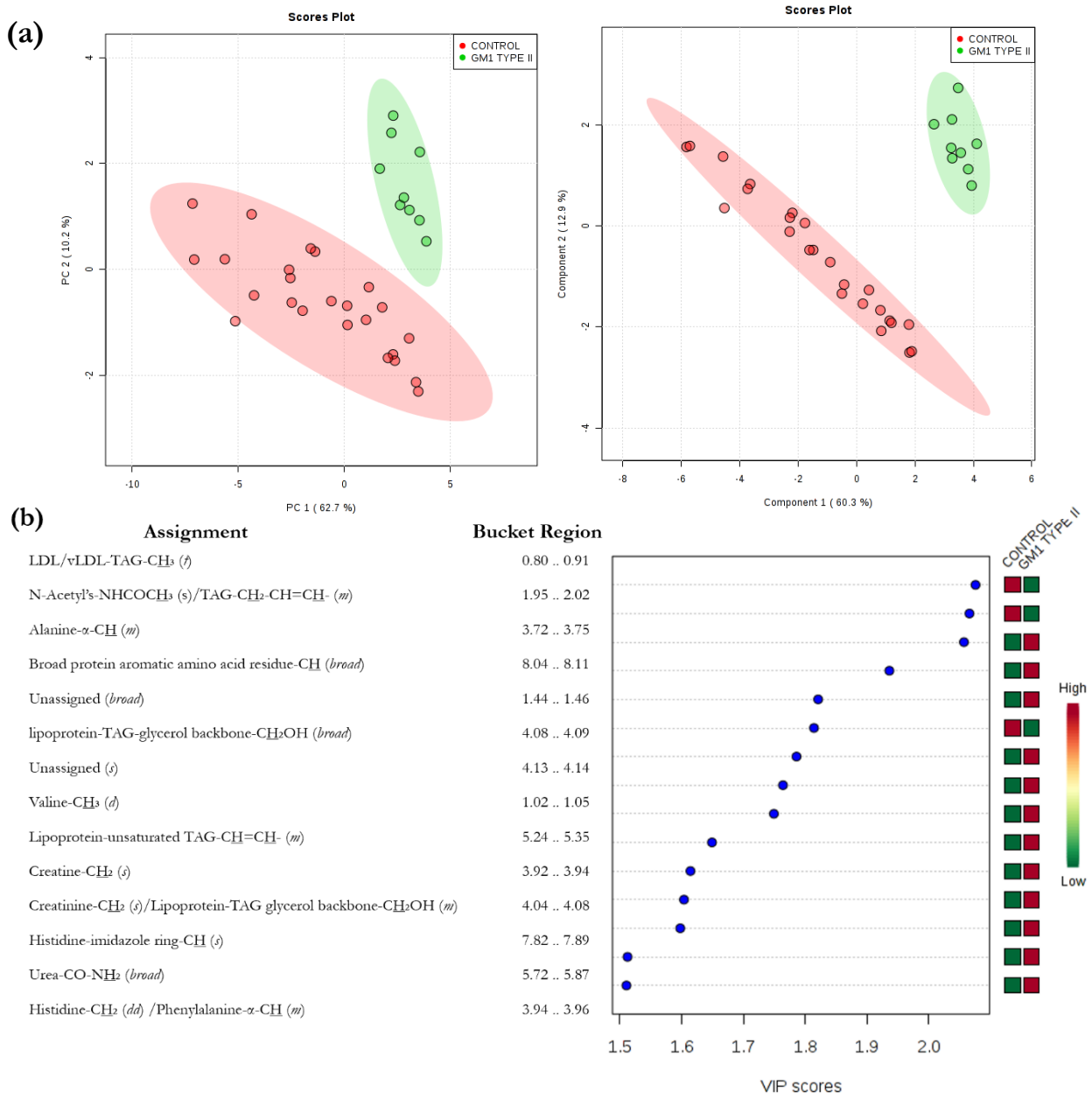


Figure 5.4. (a) Left: PCA scores plot showing controls (red) and GM1 Type II (green) study participants, together with 95% confidence ellipsoids. PC1 and PC2 represent 62.7 and 10.2% of the total dataset variance respectively. (a) Right: PLS-DA scores plot showing controls (red) and GM1 Type II (green), along with 95% confidence ellipsoids. PC1 and PC2 represent 60.3 and 12.9% of the total dataset variance respectively. Removal of all glucose bucket regions, which were clearly elevated in GM1 Type II participants in view of non-fasting of participants, was performed prior to performing this analysis. **(b)** VIP scores plot attributable to the PLS-DA analysis showing significantly elevated bucket regions, and corresponding metabolite assignments. Red and green represent metabolites elevated or depleted in GM1 Type II and controls respectively. The higher the VIP score, the more important the bucket region is considered in this PLS-DA model (all VIP values ≥ 1 were

considered significant). Abbreviations: LDL, Low density lipoprotein; TAG, triacylglycerol; vLDL, very low-density lipoprotein.

Using multivariate statistical testing, a clear distinction was found between the healthy controls and GM1 Type II plasma profiles. There was an outlier in the control group in the PCA plot that was removed prior to the final PCA and PLS-DA analysis for the reasons provided in Chapter 4.

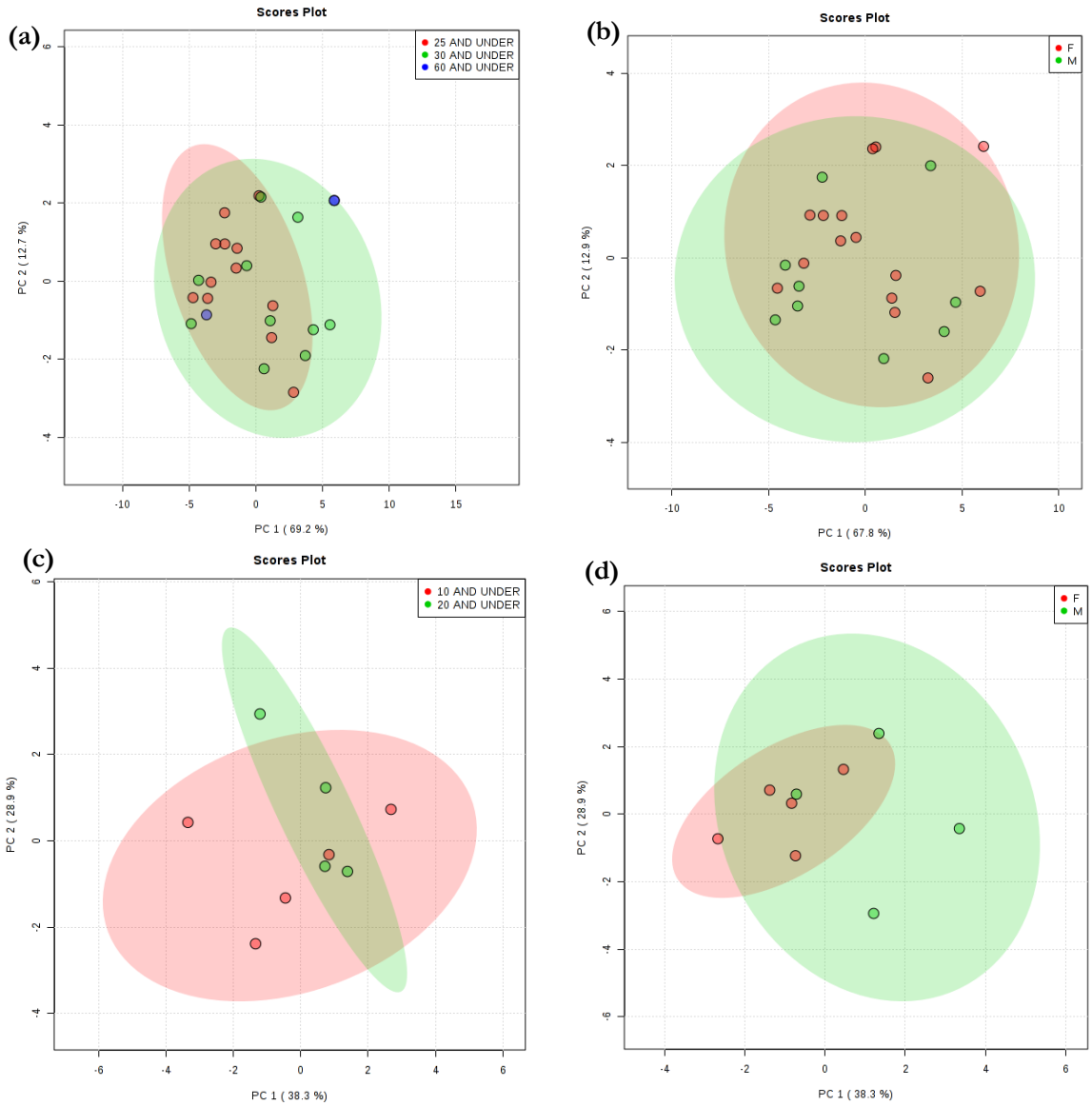


Figure 5.5. (a) PCA scores plot of plasma profiles from controls grouped by age 25 years or under, 30 years or under and 60 years or under, shown in red, green and blue, respectively. PC1 and PC2 are shown representing 69.2 and 12.7% of the dataset variance respectively. 95% confidence ellipses are shown. PLS-DA analysis was also performed on this dataset (data not shown), and the highest Q^2 value obtained was only 0.22 on a 3 component model using LOOCV and upon permutation testing using separation by distance p = 0.43 was determined with 2,000 permutations **(b)** PCA plot of plasma profiles from controls grouped by male (M) and female (F) represented in green and red respectively. PC1 and PC2 are shown representing 67.8 and 12.9% of the dataset variance respectively. 95% confidence

ellipses are shown. PLS-DA analysis was also performed on this dataset (data not shown), and the highest Q^2 value obtained was only -0.35 on a 6 component model using LOOCV and upon permutation testing using separation by distance, a p value of 0.388 was determined with 2,000 permutations. **(c)** PCA scores plot of plasma profiles from GM1 Type II patients grouped by age 10 years or under and 20 years or under shown in red and green, respectively. PC1 and PC2 are shown representing 38.3 and 28.9% of the dataset variance respectively. 95% confidence ellipses are shown. PLS-DA analysis was also performed on this dataset (data not shown) and the highest Q^2 value obtained was only 0.24 on a 5 component model using LOOCV, and upon permutation testing using separation by distance $p = 1$ was determined using 2,000 permutations. **(d)** PCA scores plot of plasma profiles from GM1 Type II patients grouped by male (M) or female (F) or under shown in green and red, respectively. PC1 and PC2 are shown representing 38.3 and 28.9% of the dataset variance respectively. 95% confidence ellipses are shown. PLS-DA analysis was also performed on this dataset (data not shown) and the highest Q^2 value obtained was only 0.02 on a 1 component model using LOOCV and upon permutation testing using separation by distance $p = 0.582$ was determined using 2,000 permutations.

Univariate Statistical Analysis

Student's t-test analysis of the plasma from GM1 Type II and controls using FDR correction of 0.05 showed 25 metabolites as significant ($p < 0.005$), including 0.80 – 0.91 ppm (↓ very-low-density-lipoprotein (vLDL)/low-density-lipoprotein (LDL)-triacylglycerol (TAG)-C$\underline{H}_3$); 0.94 – 0.97 ppm (↑ leucine-C$\underline{H}_3$); 0.97 – 0.99 ppm (↑ valine-C$\underline{H}_3$); 1.02 – 1.05 ppm (↑ valine-CH_3); 1.34 – 1.38 ppm (↓ vLDL/LDL-bulk chain-(C$\underline{H}_2$-)$_n$); 1.38 – 1.40 ppm (↓ unassigned); 1.40 – 1.42 ppm (↓ unassigned); 1.46 – 1.50 ppm (↓ alanine-C$\underline{H}_3$); 1.54 – 1.61 ppm (↓ TAG–C$\underline{H}_2$-CH_2CO); 1.95 – 2.02 ppm (↓ TAG-C$\underline{H}_2$-CH=CH /N-Acetyl-C$\underline{H}_3$); 2.05 – 2.10 ppm (↓ N-Acetyl-CH_3); 2.40 – 2.49 ppm (↑ glutamine-C$\underline{H}_2$); 2.69 – 2.72 ppm (↑ TAG-CH=CH-C$\underline{H}_2$-CH=CH;, 2.94 – 2.97 ppm (↓ trimethylamine-C$\underline{H}_3$); 3.03 – 3.04 ppm (↑ creatinine-/creatine-N-C$\underline{H}_3$); 3.26 – 3.28 ppm (↑ trimethylamine-N-oxide-C$\underline{H}_3$), 3.59 – 3.62 ppm; (↑ unassigned), 3.72 – 3.75 ppm (↑ unassigned); 3.65 – 3.66 ppm (↓ unassigned); 3.66 – 3.68 ppm (↑ unassigned); 3.92 – 3.94 ppm (↑ tyrosine-α-C$\underline{H}$/creatine-C$\underline{H}_2$); 3.96 – 4.00 ppm (↑ histidine-α-C$\underline{H}$ /phenylalanine-α-C$\underline{H}$/creatine-C$\underline{H}_2$), 4.08 – 4.09 ppm (↑ creatinine-C$\underline{H}_2$/3-hydroxybutyrate-C$\underline{H}$) and 5.24 – 5.35 ppm (↑ TAG-C$\underline{H}$=C$\underline{H}$-).

Volcano plots and fold-change analyses identified the same 6 bucket regions as significant including 4.08 – 4.09 ppm (↑ creatinine-C$\underline{H}_2$/3-hydroxybutyrate-C$\underline{H}$); 4.04 – 4.08 ppm (↑ creatine/glycerol backbone of lipids -C$\underline{H}_2$OCOR); 8.04 – 8.11 ppm (↑ trigonelline-C$\underline{H}$); 3.94 – 3.96 ppm (↑ histidine-C$\underline{H}$); 3.26 – 3.28 ppm (↑ unassigned); and 3.92 – 3.94 ppm (↑ creatine-C$\underline{H}_2$) in GM1 Type II plasma samples.

5.10 Statistical Analysis of Urinary Dataset

Multivariate Statistical Analysis

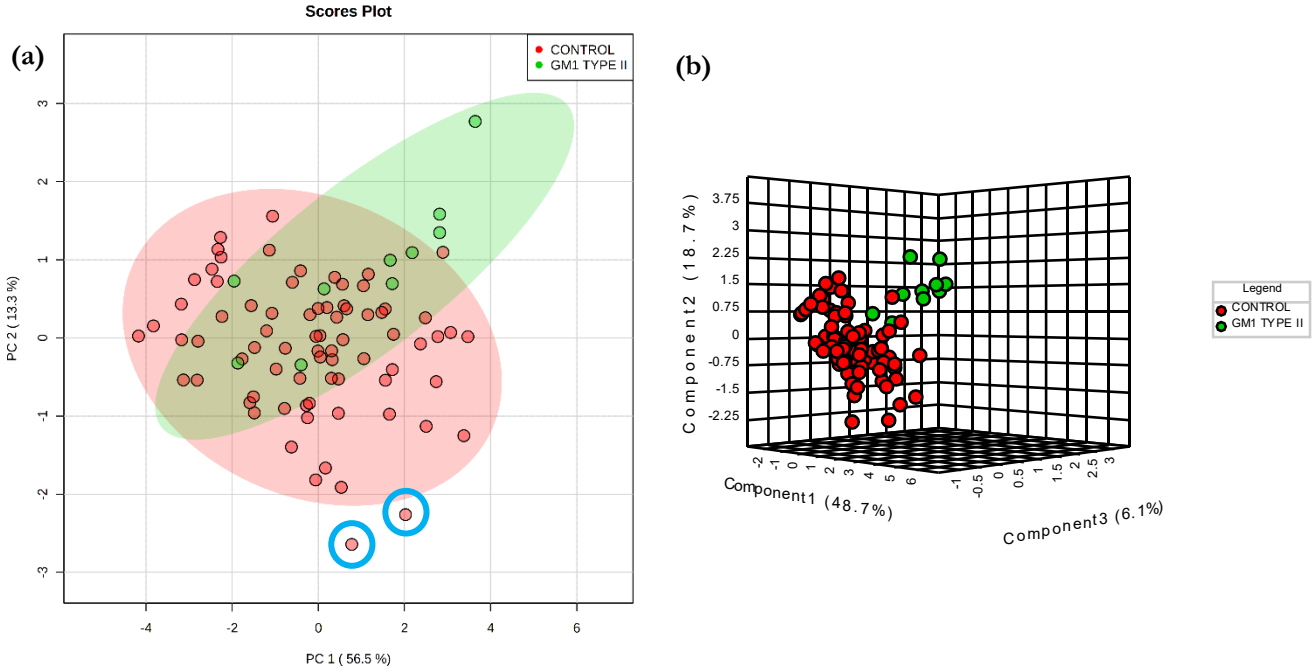


Figure 5.6. (a) 2D PCA scores plot from the ^{1}H NMR urinary profiles of controls (red) and GM1 Type II (green) participants. PC1 and PC2 are shown to represent 56.6 and 13.3% of the total dataset variance respectively. 95% confidence ellipses are shown. Outliers are circled in blue. **(b)** Corresponding 3D PLS-DA plot with control (red) and GM1 Type II (green) participants denoted. PC1, PC2 and PC3 are shown to represent 48.7, 18.7 and 6.1% of the total dataset variance respectively. Accuracy = 0.965; $Q^2 = 0.56$; and $R^2 = 0.73$ using LOOCV on a model with 4 components. Permutation testing obtained: $p < 0.0005$, using 2,000 permutations.

Bucket Region (δ/ppm)	Upregulated or downregulated in GM1 Type II	Assignment
2.52-2.59	↑	Citrate-C<u>H</u>$_{2A/B}$
3.56-3.58	↑	Glycine-C<u>H</u>$_2$
2.66-2.73	↑	Citrate-C<u>H</u>$_{2A/B}$/DMA-C<u>H</u>$_3$
2.05-2.08	↑	N-acetyl-C<u>H</u>$_3$
2.34-2.35	↑	Pyruvate-C<u>H</u>$_3$
3.23-3.24	↑	TMAO-C<u>H</u>$_3$
1.33-1.34	↑	Lactate-C<u>H</u>$_3$
2.25-2.28	↑	Acetoacetate-C<u>H</u>$_3$
1.47-1.50	↑	Alanine-C<u>H</u>$_3$
2.02-2.05	↑	N-Acetyl-C<u>H</u>$_3$
1.27-1.28	↑	Acetoacetate-C<u>H</u>$_3$
1.92-1.93	↑	Lysine-C<u>H</u>$_2$/Acetate-C<u>H</u>$_3$
4.14-.4.18	↑	Lactate-C<u>H</u>
4.29-4.31	↑	Threonine-γ-C<u>H</u>

1.93-1.96	↑	Lysine-C<u>H</u>$_2$

Table 5.2. Top significant bucket regions and corresponding metabolites that are upregulated or downregulated in GM1 Type II urinary profiles. These results are derived from PLS-DA plot in Figure 5.4. VIP scores were all ≥ 1, and therefore considered significant.

MV testing revealed some differences in PCA analysis; however, complete separation between the two groups was not possible. GM1 Type II urinary profiles loaded more positively on PC1 and PC2 than control patients. Loadings on the PCA were investigated, and both the 3.03 – 3.08 and 4.06 – 4.12 ppm bucket regions loaded negatively on PC1 and positively on PC2. These bucket regions are attributable to creatinine/creatine-N-C<u>H</u>$_3$ and creatinine-CH_2 respectively. Outliers were identified and circled in blue, and hence these samples were removed from further statistical analysis. PLS-DA was able to successfully distinguish between GM1 Type II and control participants urinary profiles (Figure 5.4). The

VIP scores plot revealed many significant metabolites (Table 5.1). The model was accurate and performed well under permutation testing (Figure 5.5).

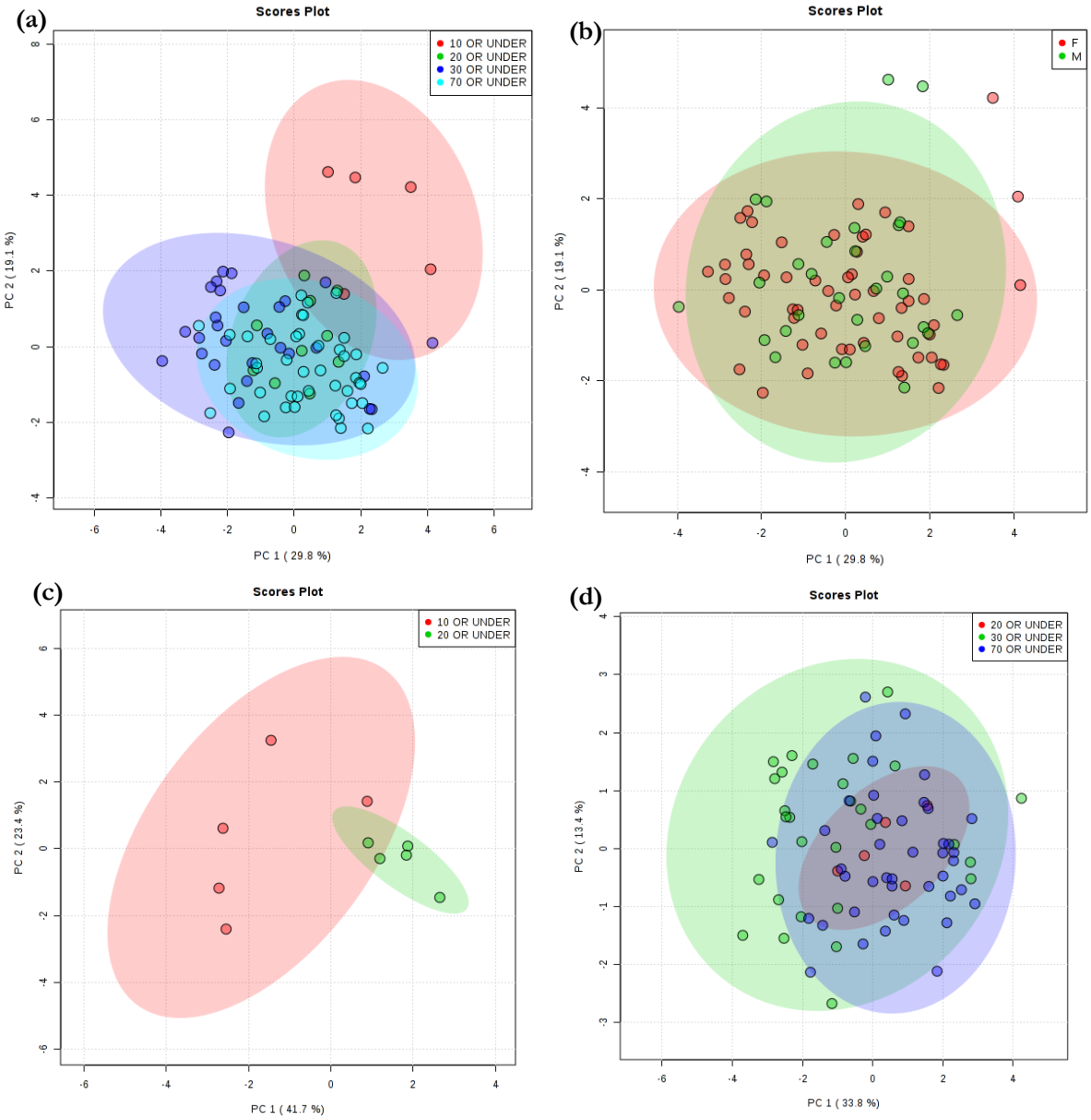


Figure 5.7. (a) PCA scores plot of urinary profiles from controls and GM1 Type II grouped by age 10 years or under, 20 years or under, 30 years or under and 70 years or under, represented in red, green, blue and turquoise, respectively. PC1 and PC2 are shown representing 29.8% and 19.1% of the dataset variance respectively. 95% confidence ellipses are shown. **(b)** PCA plot of urinary profiles from controls and GM1 Type II grouped by male and female, represented in green and red, respectively. PC1 and PC2 are shown representing 29.8% and 19.1% of the dataset variance respectively. 95% confidence ellipses are shown. **(c)** PCA scores plot of GM1 Type II samples only grouped by age 10 years or under and 20 years or under in red and green respectively. PC1 and PC2 represent 41.7 and 23.4% of the dataset variance respectively. 95% confidence ellipses are shown. PLS-DA plot of GM1 Type II samples only grouped by age 10 years or under and 20 years and under in red and green respectively was subjected to overfitting (Figure not shown here). PC1 and PC2 represented 41.1 and 17.1% of the dataset variance respectively and the highest Q^2 value obtained was 0.63 using a 1 component model using leave one out cross validation.

Permutation testing using 2000 permutations determined a p value of $p = 0.2445$ showing overfitting as a result of small sample numbers. **(d)** PCA scores plot of control samples only grouped by age 20 years or under, 30 years or under and 70 years or younger in red, green and blue respectively. PC1 and PC2 represent 33.8 and 13.4% of the dataset variance respectively. 95% confidence ellipses are shown.

Univariate Statistical Analysis

Not only were there significant differences found in MV testing, but also in UV testing. Student's t-tests with an FDR cut off value of $p < 0.05$ found the 1.07 – 1.09 ppm, 4.06 – 4.12 ppm and 9.11 – 9.14 ppm frequency buckets to be significant. These metabolites are attributable to upregulated methyl succinate-C$\underline{H}_3$ and downregulated creatinine-C$\underline{H}_2$ and trigonelline-C2-CH resonances respectively in GM1 Type II patients. Further significance was found in assigned bucket regions with an FDR cut off value of $p < 0.005$, including 1.27 – 1.28 ppm (acetoacetate-C$\underline{H}_3$); 1.33 – 1.34 ppm (lactate-C$\underline{H}_3$); 1.36 – 1.37 ppm (2-hydroxybutyrate-C$\underline{H}_3$); 1.47 – 1.50 ppm (alanine-C$\underline{H}_3$); 1.93 – 1.96 ppm (lysine-β-C$\underline{H}_2$); 2.02 – 2.05 ppm (N-acetyl-C$\underline{H}_3$); 2.25 – 2.28 ppm (2-Hydroxyglutarate-δ-C$\underline{H}_2$); 2.34 – 2.35 ppm (pyruvate-C$\underline{H}_3$); 2.52 – 2.59 ppm (citrate-C$\underline{H}_2$); 2.66 – 2.73 ppm (citrate-C$\underline{H}_2$/DMA-CH_3); 3.03 – 3.08 ppm (creatinine/creatine-C$\underline{H}_2$); 3.56 ppm – 3.58 ppm (glycine-C$\underline{H}_2$); and 4.14 – 4.18 ppm (lactate-C$\underline{H}$). These bucket regions were all upregulated in GM1 Type II urine, with the exception of 3.03 – 3.08 ppm one which was downregulated. Heatmaps were unable to fully distinguish between GM1 Type II and control urinary profiles, and the GM1 Type II participants were clustered into two groups amongst the control participants. No trend or explanation was determined for these two groups on comparison of further possible explanatory variables, including age, gender and drug treatments.

Enrichment Analysis

Location-based enrichment analysis was performed on the urinary dataset. The lysosome was considered the most significant location in enrichment analysis ($p < 0.00005$) in view of the elevation of glycine and N-acetyls function resonances in GM1 Type II participants over those of control urinary profiles. Other significant locations revealed by enrichment analysis, in order of the most significance to first, were the peroxisome, bladder, placenta, mitochondria, spleen, pancreas and intestine with the significance of all these locations being $p < 0.00005$. Both the peroxisome and mitochondria are in close proximity to the lysosome, and hence this is expected since other cellular organelles are significant, since a deficient lysosome will impact the rest of the cell. The spleen is impacted in GM1 Type I gangliosidosis

in the form of splenomegaly, but this is not a common symptom of GM1 Type II gangliosidosis. It is, however, surprising that the bladder, intestines and placenta are significant locations in this analysis.

Diseased-based enrichment analysis was also performed on the urinary dataset. This analysis revealed carbamoyl phosphate synthetase deficiency, iminoglycinuria, sarcosinemia, citrullinemia type I, glucoglycinuria, hydroxyprolinemia (type I-II), juvenile myoclonic epilepsy, methylmalonic aciduria (mma and cbla Type) and newborns, premature infants, transient immaturity of transport enzymes as the most significant diseases correlated to the urinary profiles in GM1 Type II *versus* controls ($p<0.0000005$). All of these disease statuses are significant in this form of enrichment analysis in view elevations in glycine, and notably all these diseases have similar symptoms to those of GM1 Type II. Typical symptoms of carbamoyl phosphate synthetase deficiency, citrullinemia type I and methylmalonic aciduria are seizures which are common in GM1 Type II patients. Other symptoms that align with GM1 Type II disease include mental retardation which is also present in iminoglycinuria, glucoglycinuria and hydroxyprolinemia. Sarcosinemia has many symptoms that align with GM1 Type II patients, including intellectual disability, neurological problems, skeletal abnormalities and enlarged liver.

5.11 Discussion

Potential Disease Biomarkers in Plasma, Urine and CSF

Statistically significant differences in between the plasma and urine profiles was observed in GM1 Type II patients and those of control participants. These significant differences are highlighted in both UV and MV analysis conducted on the dataset acquired, and these are observed in Figure 5.4 and Figure 5.6; for plasma and urine; respectively. Unfortunately, matched control samples for the CSF were not obtained, in view of ethical reasons and also because of the difficulties associated with the collection of this type of sample. Hence it was not possible to perform statistical comparisons with the ^{1}H NMR dataset acquired on CSF samples. Therefore, potential biomarkers in all other available biofluids were investigated.

Plasma MV metabolomics analysis of these ^{1}H NMR datasets by PCA, RF and PLS-DA approaches revealed a clear distinction between the two disease classification groups (i.e. GM1 type II gangliosidosis *vs.* healthy controls, Figure 5.4). With PLS-DA, permutation testing demonstrated a high level of discrimination between these two classifications (p = 0.0015). VIP values also confirmed a high level of significant differences between many of

the ^{1}H NMR bucket variables; for example, downregulated lipoprotein-associated TAGs in the GM1 type II gangliosidosis patients, observations which indicate overall perturbances in lipid metabolism in GM1 Type II (Figure 5.4). This observation is of particular interest, since the GM1 Type II participants were not fasted whilst the control ones were, and therefore an elevation in lipids would be expected in the GM1 Type II participants, and not a depletion. The top five most significant metabolites in the VIP bucket regions included 0.80-0.91 (vLDL-/LDL-TAG-C$\underline{H}_3$), 1.95 – 2.02 (N-Acetyl-CH_3/TAG-C$\underline{H}_2$-CH=CH- signal), 3.72 – 3.75 (unassigned), 8.04 – 8.11 (aromatic amino acid) and 1.44 – 1.46 ppm (unassigned (*d*)) resonances (Figure 5.4). The most significant bucket found from this MV analysis was attributable to vLDL-/LDL-TAGs. Shared bucket regions attributable to TAGs have been observed to be either upregulated and downregulated, and thus it is difficult to distinguish whether the disturbance in TAGs is elevated or depleted. TAGs have been shown to be elevated in cardiovascular diseases [240], but this is not relevant to GM1 Type II disease since this condition is not the cause of mortality in these patients. More interestingly, depleted LDL and elevated TAGs have been observed in the plasma of patients with other lysosomal storage disorders, including Niemann Pick Disease Type C1 (NPC1) [26]. Overall, these disturbances are likely to be a result of changes in lipid trafficking, which is common in such disorders.

Plasma creatinine is upregulated in GM1 Type II patients in view of muscle weakness/hypotonia observed in them, or alternatively, dietary sources [148]. Elevated blood histidine has been correlated with ataxia, psychosis and tremors [148], the former of which is consistent with GM1 Type II pathology. Disturbances of the valine, isoleucine and leucine pathway in serum has also been noted in another lysosomal storage disorder, NPC1 [26]. Disturbances in branched-chain amino acids (BCAAs) in plasma could be a reflection of liver condition as this has been observed elsewhere in patients with chronic active hepatitis [142]. Both the liver and spleen have been observed as enlarged in GM1 Type II patients, and this could correlate with the previous hypothesis presented by researchers working on NPC1 disease [26].

Gangliosides in CSF, urine and plasma were not identified or quantified using ^{1}H NMR analysis, and this is ascribable to the limited sensitivity of this method. LC-MS analysis was also performed in order to complete targeted analysis to identify GM1 ganglioside in the biofluids (data not shown). However, gangliosides were not observed *in vivo* using the developed LC-MS methodology. More sensitive approaches, such as the use of LC-MS/MS,

could be applied in order to monitor and distinguish such saccharides [48]. Nevertheless, a significant increase in the N-acetyl function bucket region (2.05 – 2.08 ppm) was observed in spectra of GM1 Type II patient urine samples. This elevation in N-acetyl groups in urine could arise from other ganglioside oligosaccharides, or possibly other N-acetyl storage compounds such as N-acetylamino acids. The findings supports the work previously performed, and demonstrates that there is potential for ^{1}H NMR analysis to distinguish between control and GM1 Type II patients using NMR-based cryoprobe technology [241]. This could lead to new biomarkers for GM1 Type II disease, with a combination of other markers. This may not be a biomarker specifically for GM1 Type II, since N-acetyl function upregulations have been correlated previously with many other lysosomal storage disorders, and therefore serves as a general indicator of lysosome metabolism imbalance [241]. The future employment of N-acetyl storage compounds as potential biomarkers in urine would also require further validation.

Upregulated urinary and plasma lactate have been associated with many different diseases, for example, type I diabetes [242], ethanol intoxication [243], liver failure [244] and colorectal cancer [245], and is considered an oncometabolite, an immunosuppressant, an acidogen, and a metabotoxin [148]. Indeed, elevated urinary lactate in GM1 Type II patients, which in combination with other markers, could be considered as another unspecific marker for GM1 Type II disease. Elevated urinary pyruvate and citrate may, however, indicate disturbances of the TCA cycle. As previously explored, urinary glycine levels were also shown to be extremely significant in enrichment analysis, and its upregulation may be associated with seizures. Overall, there a combination of potential non-specific biomarkers that could be used for GM1 Type II gangliosidosis diagnosis, and for the potential assessment of disease severity.

Xenobiotics/Drugs

Few metabolomics studies have been performed which assess the effect of the drugs on the metabolome as previously explored in Chapter 3. Potentially, drugs could be influencing biofluid metabolite levels; however, extensive studies addressing the effect of particular drugs on the human urine and plasma metabolome have yet to be performed. One of the participants were receiving no therapy, but the remainder were on a combination of medications. As required, none of the control participants were taking any medications or supplements.

Miglustat (n = 1) and valproate (n = 1) have both been previous explored in the urinary profiles in patients with NPC1 disease [201], and the valproate glucuronide metabolite and miglustat were also observed in urine collected from such GM1 Type II patients (Figure 5.3). Typically, these spectral regions must be removed from the spectrum prior to a metabolomics investigation. This hinders the experiment, since many metabolites can no longer be considered as part of the investigation. But with so few GM1 Type II samples, each with different drugs, it is a complex procedure to remove all metabolites associated with drugs, and their fasting and age status, using the described ^{1}H NMR strategy.

Clonidine (n = 3) is often received after opioid addiction or to help with withdrawal. Few metabolomics studies have been performed on urine and plasma prior and post taking this drug; however, rat brains have been analysed by NMR described elsewhere [246]. On this treatment glutamine, glutamate, creatine and taurine were demonstrated to return to normal in the brain with treatment of clonidine [246]. Therefore, it could be hypothesised that clonidine has capability to change metabolites in the blood plasma and even urinary metabolites, although metabolomics studies are yet to be performed to support this hypothesis. On multivariate testing, these samples (n = 3) did not appear to cluster together away from the other samples, an observation which may suggest no metabolic changes in the urine or plasma are observed, but this is only with very few numbers of samples, and a full pharmacodynamics investigation would therefore be required in order to justify this. Clonidine is usually eliminated unchanged by urinary excretion, and therefore it is surprising no signals attributable to this drug are observed. However, this most likely arises from a lengthy period between the sample being collected and dose administration, or the concentration was too low to be observed by ^{1}H NMR analysis.

Lamictal glycerine suppositories (n = 1) and Lamictal (n = 1) are the brand names for lamotrigine, and are taken by GM1 Type II patients on medication for seizures. Interestingly, 10% of patients taking that lamotrigine have skin rashes as a side-effect. The epidermis was one of the locations of interest in the GM1 Type II group. However, similarly to the patients on clonidine, these patients did not appear to cluster together on multivariate analysis using PLS-DA/PCA or appear as outliers in the analysis. Resonances related to drugs were identified. However, they did not correspond to lamotrigine or corresponding metabolites identified in the literature in urine samples, and therefore a full elucidation of the structural nature of the drugs/corresponding metabolites was not completed (Figure 5.3). To date, this is the first time that this lamotrigine /metabolite(s) has be observed in urinary profiles by ^{1}H

NMR, and 2D NMR spectroscopic analysis would be required in order to confirm their structural nature.

Tolperisone (n = 1) is a muscle relaxant taken by a patient in the GM1 Type II cohort. To date, no metabolomics investigations have been performed on tolperisone in urine and plasma so the impact of the participant taking this drug on this trial is not fully known. The patient has not appeared as an outlier in the study from the multivariate analysis. No any drug signals were assigned as a result of receiving this medication in the GM1 Type II patient. Tolperisone has been demonstrated to be excreted in urine, but only less than 0.1% of the dose was detected [247].

Levetiracetam (n = 1) is used to treat epilepsy. For the first time, signals attributable to levetiracetam and/or a related metabolite in CSF have been assigned (Figure 5.0(a)). 2D NMR analysis would be required in order to confirm the metabolite signals. The predicted NMR structure of levetiracetam is available on the HMDB as a solution in $CDCl_3$ only. From this, it was possible to locate signals that were attributable to the same patterns in the urinary profile acquired. This drug was not observed in blood plasma, only in the CSF and urine.

Albuterol (n = 1) is used to treat asthma and was taken by a patient in the GM1 Type II cohort. In a previous metabolomics study, observing the changes in plasma in autistic children, the participant receiving this drug did not appear as an outlier, which shows promise for this metabolomics study in that it may not impact metabolites in plasma[248], although urine and CSF remain unexplored. Notwithstanding, the above study used a more sensitive MS methodology [248] and not an NMR-based investigation as performed here.

Cyclopentolate (n = 1) is a muscarinic antagonist, and was used to treat a patient in the GM1 Type II cohort. To date, no metabolomics investigations have been performed on cyclopentolate in urine and plasma so the impact of the participant taking this drug on this trial is not fully known. The patient did not appear as an outlier in the study in the multivariate analysis performed. No drug or drug-related ^{1}H NMR signals in spectra acquired on this GM1 Type II patient. Since this drug is prescribed as an eye drop, it is perceived that it is unlikely to have much impact on the urinary or plasma profiles.

A GM1 Type II patient was prescribed diazepam (n = 1). The patient was 3 years old meaning a child dosage was formulated and received which is maximum 10 mg. A metabolomic study has been performed on diazepam however; it concerned the brain and not urine or plasma like in these studies [249]. No resonances associated with diazepam were observed and this

might be due to the small dosage, mass spectrometry has been the main technique of focus for detection in biofluids in view of sensitivity issues [250].

It has been previously highlighted that drugs can be detected in ^{1}H NMR spectra and impact the overall statistical analysis results acquired, and therefore, it is important to take into account where drugs are resonating and either remove them entirely from the spectra, remove the sample or handle the significance of the results with caution. These buckets were removed prior to the performance of multivariate analysis as described in previous chapters.

Drug Metabolite	Chemical Shift (δ/ppm)	Urinary Metabolite(s) that appear in region	Significant buckets in GM1 Type II
1-O-valproyl-β-glucuronide	0.88		No
	1.31	Lactate-CH_3	No
	1.53		No
	1.61	L-Lysine-CH_2	No
	2.59	Citrate-CH_2	Yes
	3.86	Mannitol-CH_2	No
	3.52	Phenylacetylglutamine-CH_3	
	3.59		No
Miglustat	1.02	Valine-CH_3	No
	1.45	Lysine-CH_2	No
	1.74		No
	2.99		No
	3.05	Taurine-CH_2	No
	3.20	Methanol-CH_3	No
	3.39		No
	3.45	Glycine-CH_2	No
	3.60	Lysine-CH_2	No
	3.70	Creatinine-CH	No
	3.91		No
	4.12		No
Levetiracetam	1.67-1.72	Lysine	No
	2.50-2.56	Citrate-$CH_{2A/B}$	No
	4.44-4.48	Trigonelline	No
Lamotrigine/metabolite	0.86		No
	1.56		No
	1.73		No
	2.59		No

Table 5.3. Metabolite signals impacted by drug metabolism of the various medications which were removed prior to the multivariate and univariate analysis conducted.

Diet and Dietary Supplements

Omega-3 PUFA (n = 1) and vitamin D3 (n = 1) that have been taken by GM1 Type II patients have already been previously explored in Chapter 3 in addition to dietary

supplements. However, vitamin B6 (n = 1) was not investigated as part of this discussion. It has been found that vitamin B6 increases the ratios of glutamine/glutamate and 2-oxoglutarate/glutamate in plasma alongside upregulations in acetate, pyruvate and TMAO [251]. As only one participant was taking vitamin B6 as a supplement it is unlikely that this would impact the statistical analysis of the whole cohort and this participant did not appear as an outlier as a result of taking this vitamin. No resonances in the ^{1}H NMR spectra from administration of this vitamin were observed, largely because the dose applied was too small for it to be ^{1}H NMR-detectable.

Fasting and Non-fasting status

Many of the metabolites can be introduced as a result of dietary differences between the fasted and unfasted cohort. This has been explored extensively in Chapter 3, but the metabolites that are significant in this particular investigation, and which may be as ascribable to a dietary difference, include citrate and amino acids, for example. The issues of analysis of fasted and unfasted samples have been previously addressed in Chapter 3, and similar considerations arise within this study with regard to specific metabolites such as citrate and amino acids, which can be ingested with corresponding increases in blood plasma and urine accordingly. Histidine, trigonelline and valine concentrations were also elevated in GM1 Type II patients, but this is likely to arise from dietary sources, and the different fasting status of the participants involved in this investigation [148]. Additionally, disturbances in TAGs in GM1 Type II when expressed relative to control plasma samples were observed, which could be attributable to diet; this has been demonstrated elsewhere [252], [253].

Sample Volume and Sample Storage

The volume of sample was limited in this study, therefore less volume was used for analysis than that in the investigations described in previous chapters; a higher quantity of urine and plasma may reveal more sensitive results and potentially more metabolites in which are only present at lower concentrations; however, the results acquired are consistent with molecules previously identified in ^{1}H NMR studies of CSF [63]. Some methodologies have employed freeze-drying techniques using CSF to increase metabolite concentrations; however, this is usually useful for larger volumes of CSF. Sensitivity issues were avoided by using the highly sensitive 700 MHz NMR at Oxford University, and to date gangliosides have not been identified in biofluids *via* ^{1}H NMR or LC-MS *in vivo*, which may have served as useful biomarkers.

Sample storage length could also impact on metabolite concentrations in the samples investigated, and this has been demonstrated elsewhere [102]. Samples analysed over a 2-year period demonstrated that metabolite determinations in the blood serum of fasting individuals are more reproducible, than those of unfasted individual sera [102]. This factor could also partially influence the results acquired here, since the GM1 type II samples were collected in the 2011-2017 period, and hence were stored for longer than the control samples, which were all collected in 2017. This presents some challenges in sample collection over a longitudinal period, which is essential in assessing disease severity over a period of time in rare disorders.

Age and Sample Numbers

Another limitation of this study is the limited number of samples available in this study. This is ascribable to the rarity of the disease studied. Furthermore, age- matched controls were not collected in view of the ethical approval for them only being valid for participants aged over 18 years. The importance of age-matching is, however, quite vital in metabolomics studies, since metabolic patterns modify with age. For example, it has been demonstrated that in the ^{1}H NMR profiles of urine analysed in children of the ages of 12 and below that, creatine, creatinine, glycine, betaine/TMAO, citrate, succinate and acetone were all correlated with age [254]. Creatinine was the only metabolite that elevated with age, whereas that of all others decreased [254]. Creatinine and creatine are metabolites that are impacted by age, and are no doubt significant in this study in view of the different age ranges in the cohorts studied, and this has been demonstrated as a contributing metabolite in the PCA scores in this study. However, glycine, TMAO and citrate increased in the GM1 Type II patients' urine, which are the younger cohort.

Correlation networks have also been applied to the urinary profiles of older participant cohorts aged 40.9±10.3 years analysed by Liquid Chromatography-High Resolution Mass Spectrometry (LC-HRMS), and it was found that a number of metabolites were statistically associated with age [112]. Not only are these changes present in younger participants and urine, but on comparison of plasma profiles in elderly groups with centurions using ^{1}H NMR, the latter group were found to have significantly higher citrate, phenylalanine, creatinine, glutamine, phoshatidylcholine and N-acetylglycoproteins [255]. This correlation with age and the ^{1}H NMR profiles acquired in this study is important, since the cohorts investigated were not aged-matched, and some participants in the GM1 Type II cohort as young as 3 may have impacted on the statistical loadings in the profiles and hence impacted upon the sample

scores in PCA plots. This is one of the challenges in metabolomics; indeed, imperfect age matching renders it difficult to interpret results acquired in such inadequate study designs. In view of the rarity of GM1 Type II participants and the critical condition of these patients, clearly it is difficult to control all such variabilities.

The most significant blood plasma ^{1}H NMR frequency bucket found from this MV analysis was attributable to vLDL/LDL-TAGs; however, this may arise from age differences between GM1 Type II and the control participants, i.e. the GM1 Type II patients are younger with mean± SEM ages of 11±2.3 years versus 22±0.7 years for the non-age-matched controls in plasma profiles respectively (Table 5.0). Multivariate analysis differences observed between ages is more profound in the urine samples used in this investigation (Table 5.1). This highlights the importance of ensuring that such metabolomics studies recruit participants who are age-matched as much as possible. However, plasma creatinine levels are expected to elevate with increasing age [256], but the reversed effect was observed in our study participants, with the GM1 Type II patients representing the younger cohort.

Figures 5.5 and 5.7 show that the GM1 Type II patients and control plasma and urine samples are able to be distinguished using PCA by age. The strongest correlations are revealed in the younger age group, i.e. < 10 years of age in both plasma and urine samples. As participants increase in age, the clustering becomes more variable within the age group in both biofluids. This again emphasises the requirement for an age-matched cohort in metabolomics investigations. Figures 5.5 and 5.7 also demonstrate the impact of gender on matching on the investigation. Clustering in the urinary profiles is completely overlaid and it is difficult to observe major urinary changes related to gender using the buckets selected. Moreover, in the plasma profiles, it was observed that no clear distinction between genders were found. However, this separation could, in principle, also arise from the fasting and unfasting status of these groupings.

Unbalanced sample sizes can also lead to some unreliability regarding the predictive accuracy of the model explored [257]. This can be counteracted by using a sub-set of control samples [257]. More urine samples are present in the control group, and not many in the diseased group, this may cause some insignificant variables to appear significant.

On reperforming this same analysis with controls (n = 11, mean±SEM age of 21.5±0.74, 2 males and 9 females) and GM1 Type II patients (n = 10, ages and genders provided in Table 4.0), the statistical modelling process was improved, as expected. RF was able to distinguish between the samples with no errors or misclassifications. PCA showed no outliers, and all

samples were within the 95% confidence ellipses for each cluster. PLS-DA revealed the same metabolites as significant in the corresponding VIP plot, and performed well with an accuracy of 1.0, (R^2 = 0.92 and Q^2 = 0.757). Bucket regions with VIP scores ≥1 comprised 3.56 – 3.58 ppm (glycine-C<u>H</u>$_2$), 3.03 – 3.08 ppm (creatinine/creatine-N-C<u>H</u>$_2$), 2.05 – 2.08 ppm (N-acetyl functions-C<u>H</u>$_3$), 2.52 – 2.59 ppm (citrate-C<u>H</u>$_2$), 2.34 – 2.35 ppm (pyruvate-C<u>H</u>$_3$), 1.33 – 1.34 ppm (lactate-C<u>H</u>$_3$), 4.06 – 4.12 ppm (creatinine-C<u>H</u>$_2$), 1.92 – 1.93 ppm (acetate-CH$_3$/lysine-β-C<u>H</u>$_2$), 2.25 – 2.28 ppm (2-hydroxyglutarate-δ-C<u>H</u>$_2$), 2.02 – 2.05 ppm (N-Acetyl functions-C<u>H</u>$_3$), 1.47 – 1.50 ppm (alanine-C<u>H</u>$_3$), 1.93 – 1.96 ppm (lysine-β-C<u>H</u>$_2$) and 1.36 – 1.37 ppm (2-hydroxybutyrate-C<u>H</u>$_3$). SVM was able to distinguish fully between the two groups with 0% error with 41 variables, and 11.1% with only 25 variables.

Univariate statistical analysis revealed slightly different results as three bucket regions had a fold-change of > 2.0, and these were 2.78 – 2.79 ppm (dimethylamine-N(C<u>H</u>$_3$)$_2$, 3.56 – 3.58 ppm (glycine-C<u>H</u>$_2$), and 1.92 – 1.93 ppm (acetate-C<u>H</u>$_3$/lysine-C<u>H</u>$_2$). Student's t-test revealed upregulations in 1.27 – 1.28 ppm (acetoacetate-C<u>H</u>$_3$), 1.36 – 1.37 ppm (2-hydroxybutyrate-CH$_3$), 1.47 – 1.50 ppm (alanine-C<u>H</u>$_3$), 1.92 – 1.93 ppm (acetate-CH$_3$/lysine-β-CH$_2$), 2.02 – 2.05 ppm (N-Acetyl functions-C<u>H</u>$_3$), 2.05 – 2.08 ppm (N-Acetyl functions-C<u>H</u>$_3$), 2.25 – 2.28 ppm (2-Hydroxyglutarate-δ-C<u>H</u>$_2$), 2.34 – 2.35 ppm (pyruvate-C<u>H</u>$_3$), 2.52 – 2.59 ppm (citrate-C<u>H</u>$_2$), 2.66 – 2.73 ppm (citrate-C<u>H</u>$_2$/DMA-C<u>H</u>$_3$), and 4.14 – 4.18 ppm (lactate-C<u>H</u>), along with downregulations in 4.06 – 4.12 ppm (creatinine-C<u>H</u>$_2$) in GM1 Type II of which have an FDR-corrected p value of < 0.05. Student's t-tests revealed upregulations of 1.33 – 1.34 ppm (lactate-C<u>H</u>$_3$) and 3.56 – 3.58 ppm (glycine-C<u>H</u>$_2$), and downregulations in 3.03 – 3.08 ppm (creatinine/creatine-N-C<u>H</u>$_2$) in GM1 Type II patients were also significant (p < 0.005).

5.12 Conclusions

^{1}H NMR analysis provides an insight into some new potential biomarkers in plasma and urine samples in patients with GM1 Type II disease, particularly disease relevant N-Acetyl groups, which appear as a chemopathological markers in urine, in addition to other metabolites such as lactate, citrate and pyruvate. Disturbances in plasma LDL and its associated TAGs appear to represent a common marker amongst lysosomal storage disorders. The many complications of this study were explored, including participant fasting status, drug therapies, and the genders and ages of the participants in both cohorts, and the potential impacts of these on the experimental design and implications on the results

acquired were fully investigated. The LC-MS approach applied was unable to identify gangliosides in plasma or CSF samples, although more sensitive MS strategies could be employed in a future investigation inclusive of LC-MS/MS.

Overall Thesis Conclusions

An extensive review of the literature in Chapter 1 revealed key new areas of developments in metabolomics analyses for Gaucher's Disease, GM1 Gangliosidosis and Niemann Pick Disease Type C. These included (1) systematic sample collection and handling; (2) further validation of at least some of the potential biomarkers suggested via the rigorous monitoring of their biofluid levels along with simultaneous co-ordinated assessments of clinical responses to treatment(s); (3) detailed correlations of biomarker concentrations with disease severity, and also the strict consideration of demographic variables such as age, gender and diet, etc.; and (4) a further detailed review in order to ensure that biomarker(s) identified are solely specific to particular LLSDs. This thesis aimed to contribute to developments in (3) and (4) within this field.

Chapter 2 highlighted current, appropriate methods to approach metabolomics investigations.

Chapter 3 demonstrates advancements in Gaucher's Disease metabolomics by ^{1}H NMR analysis of plasma and urine samples using a ^{1}H NMR strategy. Statistical techniques such as PLS-DA showed significant metabolic differences between GD and control cohorts of participants, and these included disturbances in plasma lipid profiles, e.g. the TAG profiles of vLDL, LDL and HDL. No statistical differences were observed on comparison of the urinary profiles of Gaucher's Disease with control participants (drugs and drug metabolite resonances were removed from spectral profiles prior to analysis). Unfortunately, Gaucher Disease patients were receiving a range of treatments, and this may complicate such metabolomics studies.

Chapter 4 investigates the ^{1}H NMR profiles of urine collected from NP-C1 diseased felines, along with the influence of hydroxypropyl-β-cyclodextrin (HPβCD) therapy on these. Statistically significant differences were found on comparison of untreated NPC felines and treated 1,000 mg SQ/IT 120 mg NPC felines. Moreover, statistically significant differences between controls and NPC feline urine profiles were only observed in older felines (age range 16-24 weeks). Notwithstanding, imperfect sex-matched cohorts involved may be contributing to the statistical differences observed, and this was further explored as part of this investigation. Potential metabolic disturbances in NPC felines included upregulations in felinine, for example.

Chapter 5 shows for the first time the ^{1}H NMR analysis of GM1 Type II Gangliosidosis CSF, plasma and urine samples, and a metabolomics investigation was performed. Upregulated citrate, N-acetyl function-containing biomolecules, creatinine, creatine and lactate, amongst other metabolites, were observed in GM1 Type II urine over those of controls. Downregulated LDL-TAGs and upregulated creatinine and lactate, were examples of alternative metabolic disturbances observed.

In conclusion, further research investigations are required in order to validate the findings observed here, which involve the careful design of such studies featuring (1) systematic sample collection and handling; and (2) further validation of at least some of the potential biomarkers indicated via the rigorous monitoring of their biofluid levels, together with simultaneous co-ordinated assessments of clinical responses to treatment(s). However, such studies remain challenging in view of the critical illness of patients with LSDs, and relatively small sample sizes in the human investigations described here.

References

[1] A. C. Dona *et al.*, "A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments," *Comput. Struct. Biotechnol. J.*, vol. 14, pp. 135–153, 2016.

[2] F. M. Platt, "Sphingolipid lysosomal storage disorders," *Nature*, vol. 510, no. 7503, pp. 68–75, 2014.

[3] F. M. Platt, A. D'Azzo, B. L. Davidson, E. F. Neufeld, and C. Tifft, "Lysosomal Storage Diseases," *Nat. Rev. Dis. Prim.*, vol. 4, no. 27, pp. 1–25, 2018.

[4] K. Abe and N. Sakai, "Patient with Niemann-Pick disease type C: Over 20 years' follow-up," *BMJ Case Rep.*, vol. 2017, 2017.

[5] L. Essabar, T. Meskini, N. Lamalmi, S. Ettair, N. Erreimi, and N. Mouane, "Gaucher's disease: Report of 11 cases with review of literature," *Pan Afr. Med. J.*, vol. 20, 2015.

[6] D. K. Trivedi, K. A. Hollywood, and R. Goodacre, "Metabolomics for the masses: The future of metabolomics in a personalized world," *New Horizons Transl. Med.*, vol. 3, no. 6, pp. 294–305, 2017.

[7] A. McCartney *et al.*, "Metabolomics in breast cancer: A decade in review," *Cancer Treatment Reviews*, vol. 67. W.B. Saunders Ltd, pp. 88–96, 01-Jun-2018.

[8] A. K. Kaushik and R. J. DeBerardinis, "Applications of metabolomics to study cancer metabolism," *Biochimica et Biophysica Acta - Reviews on Cancer*, vol. 1870, no. 1. Elsevier B.V., pp. 2–14, 01-Aug-2018.

[9] E. Saracchi, S. Fermi, and L. Brighina, "Emerging candidate biomarkers for Parkinson's disease: A review," *Aging and Disease*, vol. 5, no. 1. International Society on Aging and Disease, pp. 27–34, 2014.

[10] J. Liu *et al.*, "Ganglioside GD3 synthase (GD3S), a novel cancer drug target," *Acta Pharmaceutica Sinica B*, vol. 8, no. 5. Chinese Academy of Medical Sciences, pp. 713–720, 01-Sep-2018.

[11] M. Fan *et al.*, "Identification of Niemann-Pick C1 disease biomarkers through sphingolipid profiling.," *J. Lipid Res.*, vol. 54, no. 10, pp. 2800–14, Oct. 2013.

[12] B. C. Percival, M. Gibson, P. B. Wilson, F. M. Platt, and M. Grootveld, "Metabolomic Studies of Lipid Storage Disorders, with Special Reference to Niemann-Pick Type C Disease: A Critical Review with Future Perspectives," *Int. J. Mol. Sci.*, vol. 21, no. 7, p. 2533, Apr. 2020.

[13] V. Ruiz-Rodado *et al.*, "^{1}H NMR-Linked Urinary Metabolic Profiling of Niemann-Pick Class C1 (NPC1) Disease: Identification of Potential New Biomarkers using Correlated Component Regression (CCR) and Genetic Algorithm (GA) Analysis Strategies," *Curr. Metabolomics*, vol. 2, no. 2, pp. 88–121, 2014.

[14] B. E. Kennedy, A. S. Hundert, D. Goguen, I. C. G. Weaver, and B. Karten, "Presymptomatic Alterations in Amino Acid Metabolism and DNA Methylation in the Cerebellum of a Murine Model of Niemann-Pick Type C Disease," *Am. J. Pathol.*, vol. 186, no. 6, pp. 1582–1597, Jun. 2016.

[15] R. Sidhu *et al.*, "A validated LC-MS/MS assay for quantification of 24(S)-hydroxycholesterol in plasma and cerebrospinal fluid," *J. Lipid Res.*, vol. 56, no. 6, pp. 1222–1233, Jun. 2015.

[16] B. Tortelli *et al.*, "Cholesterol homeostatic responses provide biomarkers for monitoring treatment for the neurodegenerative disease Niemann-Pick C1 (NPC1)," *Hum. Mol. Genet.*, vol. 23, no. 22, pp. 6022–6033, Nov. 2014.

[17] F. D. Porter *et al.*, "Cholesterol oxidation products are sensitive and specific blood-based biomarkers for Niemann-Pick C1 disease," *Sci. Transl. Med.*, vol. 2, no. 56, Nov. 2010.

[18] X. Jiang *et al.*, "A sensitive and specific LC-MS/MS method for rapid diagnosis of Niemann-Pick C1 disease from human plasma.," *J. Lipid Res.*, vol. 52, no. 7, pp. 1435–45, Jul. 2011.

[19] S. Boenzi *et al.*, "A new simple and rapid LC-ESI-MS/MS method for quantification of plasma oxysterols as dimethylaminobutyrate esters. Its successful use for the diagnosis of Niemann-Pick type C disease," *Clin. Chim. Acta*, vol. 437, pp. 93–100, Nov. 2014.

[20] S. Pajares *et al.*, "Cholestane-3β,5α,6β-triol: High levels in Niemann-Pick type C, cerebrotendinous xanthomatosis, and lysosomal acid lipase deficiency," *J. Lipid Res.*, vol. 56, no. 10, pp. 1926–1935, Oct. 2015.

[21] A. K. Giese *et al.*, "A novel, highly sensitive and specific biomarker for Niemann-Pick type C1 disease," *Orphanet J. Rare Dis.*, vol. 10, no. 1, Jun. 2015.

[22] N. Lin *et al.*, "Determination of 7-ketocholesterol in plasma by lc-ms for rapid diagnosis of acid smase-deficient niemann-pick disease," *J. Lipid Res.*, vol. 55, no. 2, pp. 338–343, Feb. 2014.

[23] M. Wang *et al.*, "Plasma 7-ketocholesterol levels and the risk of incident cardiovascular events," *Heart*, vol. 103, no. 22, pp. 1788–1794, Nov. 2017.

[24] R. W. D. Welford *et al.*, "Plasma lysosphingomyelin demonstrates great potential as a diagnostic biomarker for niemann-pick disease type C in a retrospective study," *PLoS One*, vol. 9, no. 12, Dec. 2014.

[25] W. L. Chuang *et al.*, "Lyso-sphingomyelin is elevated in dried blood spots of Niemann-Pick B patients," *Mol. Genet. Metab.*, vol. 111, no. 2, pp. 209–211, 2014.

[26] F. Probert *et al.*, "NMR analysis reveals significant differences in the plasma metabolic profiles of Niemann Pick C1 patients, heterozygous carriers, and healthy controls," *Sci. Rep.*, vol. 7, no. 1, 2017.

[27] R. Mashima, M. Maekawa, A. Narita, T. Okuyama, and N. Mano, "Elevation of plasma lysosphingomyelin-509 and urinary bile acid metabolite in Niemann-Pick disease type C-affected individuals," *Mol. Genet. Metab. Reports*, vol. 15, pp. 90–95, Jun. 2018.

[28] R. Sidhu *et al.*, " N -acyl- O -phosphocholineserines: structures of a novel class of lipids that are biomarkers for Niemann-Pick C1 disease ," *J. Lipid Res.*, vol. 60, no. 8, pp. 1410–1424, Aug. 2019.

[29] M. Maekawa *et al.*, "LC/ESI-MS/MS analysis of urinary 3β-sulfooxy-7β-N-acetylglucosaminyl- 5-cholen-24-oic acid and its amides: New biomarkers for the

detection of Niemann-Pick type C disease," *Steroids*, vol. 78, no. 10, pp. 967–972, 2013.

[30] M. Maekawa *et al.*, "Focused metabolomics using liquid chromatography/electrospray ionization tandem mass spectrometry for analysis of urinary conjugated cholesterol metabolites from patients with Niemann–Pick disease type C and 3β-hydroxysteroid dehydrogenase deficiency," *Ann. Clin. Biochem.*, vol. 52, no. 5, pp. 576–587, Sep. 2015.

[31] M. Maekawa *et al.*, "Identification of Two Sulfated Cholesterol Metabolites Found in the Urine of a Patient with Niemann–Pick Disease Type C as Novel Candidate Diagnostic Markers," *Mass Spectrom.*, vol. 5, no. 2, pp. S0053–S0053, 2016.

[32] M. Maekawa *et al.*, "Diagnostic performance evaluation of sulfate-conjugated cholesterol metabolites as urinary biomarkers of Niemann–Pick disease type C," *Clin. Chim. Acta*, vol. 494, pp. 58–63, Jul. 2019.

[33] M. Maekawa *et al.*, "Development of a simultaneous analytical method for five conjugated cholesterol metabolites in urine and investigation of their performance as diagnostic markers for Niemann-Pick disease type C," 2019.

[34] F. Mazzacuva *et al.*, "Identification of novel bile acids as biomarkers for the early diagnosis of Niemann-Pick C disease," *FEBS Lett.*, vol. 590, no. 11, pp. 1651–1662, Jun. 2016.

[35] T. Geberhiwot *et al.*, "Consensus clinical management guidelines for Niemann-Pick disease type C," *Orphanet Journal of Rare Diseases*, vol. 13, no. 1. BioMed Central Ltd., 06-Apr-2018.

[36] M. Wanikawa, H. Nakamura, S. Emori, N. Hashimoto, and T. Murayama, "Accumulation of sphingomyelin in Niemann-Pick disease type C cells disrupts Rab9-dependent vesicular trafficking of cholesterol," *J. Cell. Physiol.*, Sep. 2019.

[37] F. Probert *et al.*, "Urinary excretion and metabolism of miglustat and valproate in patients with Niemann-Pick type C1 disease: One- and two-dimensional solution-state 1H NMR studies," *J. Pharm. Biomed. Anal.*, vol. 117, no. 2, pp. 276–288, 2016.

[38] A. Matencio, M. A. Alcaráz-Gómez, F. García-Carmona, B. Arias, and J. M. López-Nicolás, "Application of a simple methodology to analyze Hydroxypropyl-β-Cyclodextrin in urine using HPLC–LS in early Niemann–Pick disease type C patient," *J. Chromatogr. B*, vol. 1093–1094, pp. 47–51, Sep. 2018.

[39] J. E. M. Groener, B. J. H. M. Poorthuis, S. Kuiper, C. E. M. Hollak, and J. M. F. G. Aerts, "Plasma glucosylceramide and ceramide in type 1 Gaucher disease patients: Correlations with disease severity and response to therapeutic intervention," *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids*, vol. 1781, no. 1–2, pp. 72–78, Jan. 2008.

[40] A. Rolfs *et al.*, "Glucosylsphingosine is a highly sensitive and specific biomarker for primary diagnostic and follow-up monitoring in Gaucher disease in a non-Jewish, Caucasian cohort of Gaucher disease patients.," *PLoS One*, vol. 8, no. 11, p. e79732, 2013.

[41] N. Dekker *et al.*, "Elevated plasma glucosylsphingosine in Gaucher disease: Relation to phenotype, storage cell markers, and therapeutic response," *Blood*, vol. 118, no. 16, Oct. 2011.

[42] M. Fuller, J. Szer, S. Stark, and J. M. Fletcher, "Rapid, single-phase extraction of

glucosylsphingosine from plasma: A universal screening and monitoring tool," *Clin. Chim. Acta*, vol. 450, pp. 6–10, Oct. 2015.

[43] T. Fujiwaki, S. Yamaguchi, M. Tasaka, N. Sakura, and T. Taketomi, "Application of delayed extraction-matrix-assisted laser desorption ionization time-of-flight mass spectrometry for analysis of sphingolipids in pericardial fluid, peritoneal fluid and serum from Gaucher disease patients," *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, vol. 776, no. 1, pp. 115–123, Aug. 2002.

[44] A. Zhang, H. Sun, and X. Wang, "Saliva metabolomics opens door to biomarker discovery, disease diagnosis, and treatment," *Appl. Biochem. Biotechnol.*, vol. 168, no. 6, pp. 1718–1727, Nov. 2012.

[45] L. Ou, M. J. Przybilla, and C. B. Whitley, "Metabolomics profiling reveals profound metabolic impairments in mice and patients with Sandhoff disease," *Mol. Genet. Metab.*, vol. 126, no. 2, pp. 151–156, Feb. 2019.

[46] K. Ikeda, T. Shimizu, and R. Taguchi, "Targeted analysis of ganglioside and sulfatide molecular species by LC/ESI-MS/MS with theoretically expanded multiple reaction monitoring," *J. Lipid Res.*, vol. 49, no. 12, pp. 2678–2689, 2008.

[47] A. D. Garcia, J. L. Chavez, and Y. Mechref, "Rapid and sensitive LC-ESI-MS of gangliosides," *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, vol. 947–948, no. 1, pp. 1–7, Feb. 2014.

[48] J. Gu, C. J. Tifft, and S. J. Soldin, "Simultaneous quantification of GM1 and GM2 gangliosides by isotope dilution tandem mass spectrometry," *Clin. Biochem.*, vol. 41, no. 6, pp. 413–417, Apr. 2008.

[49] J. R. Jarnes Utz, T. Crutcher, J. Schenider, P. Sorgen, and C. B. Whitley, "Biomarkers of central nervous system inflammation in infantile and juvenile gangliosidoses," *Mol. Genet. Metab.*, vol. 114, pp. 274–280, 2015.

[50] U. F. H. Engelke *et al.*, "N-acetylated metabolites in urine: proton nuclear magnetic resonance spectroscopic study on patients with inborn errors of metabolism.," *Clin. Chem.*, vol. 50, no. 1, pp. 58–66, Jan. 2004.

[51] Y. Sandlers, "The future perspective: metabolomics in laboratory medicine for inborn errors of metabolism," *Translational Research*, vol. 189. Mosby Inc., pp. 65–75, 01-Nov-2017.

[52] J. M. Kelly, A. Bradbury, D. R. Martin, and M. E. Byrne, "Emerging therapies for neuropathic lysosomal storage disorders," *Progress in Neurobiology*, vol. 152. Elsevier Ltd, pp. 166–180, 01-May-2017.

[53] M. Grootveld and C. J. L. Silwood, "1H NMR analysis as a diagnostic probe for human saliva," *Biochemical and Biophysical Research Communications*, vol. 329, no. 1. Academic Press Inc., pp. 1–5, 01-Apr-2005.

[54] G. Polo *et al.*, "Diagnosis of sphingolipidoses: A new simultaneous measurement of lysosphingolipids by LC-MS/MS," *Clin. Chem. Lab. Med.*, vol. 55, no. 3, pp. 403–414, Mar. 2017.

[55] B. C. Percival *et al.*, "Low-field, benchtop NMR spectroscopy as a potential tool for point-of-care diagnostics of metabolic conditions: Validation, protocols and computational models," *High-Throughput*, vol. 8, no. 1, 2019.

[56] S. L. Ramsay *et al.*, "Determination of oligosaccharides and glycolipids in amniotic fluid by electrospray ionisation tandem mass spectrometry: In utero indicators of lysosomal storage diseases," *Mol. Genet. Metab.*, vol. 83, no. 3, pp. 231–238, Nov. 2004.

[57] M. Boutin and C. Auray-Blais, "Metabolomic discovery of novel urinary galabiosylceramide∈analogs as Fabry disease biomarkers," *J. Am. Soc. Mass Spectrom.*, vol. 26, no. 3, pp. 499–510, 2015.

[58] J. C. Lindon, J. K. Nicholson, and E. Holmes, *The Handbook of Metabolomics*. Elsevier, 2011.

[59] P. S. Gromski *et al.*, "A tutorial review: Metabolomics and partial least squares-discriminant analysis - a marriage of convenience or a shotgun wedding," *Anal. Chim. Acta*, vol. 879, pp. 10–23, 2015.

[60] S. Bouatra *et al.*, "The Human Urine Metabolome," *PLoS One*, vol. 8, no. 9, 2013.

[61] N. Psychogios *et al.*, "The human serum metabolome.," *PLoS One*, vol. 6, no. 2, p. e16957, Jan. 2011.

[62] J. K. Nicholson, J. C. Lindon, and E. Holmes, "'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data," *Xenobiotica*, vol. 29, no. 11, pp. 1181–1189, 1999.

[63] D. S. Wishart *et al.*, "The human cerebrospinal fluid metabolome," *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, vol. 871, no. 2, pp. 164–173, 2008.

[64] D. D. Marshall and R. Powers, "Beyond the Paradigm: Combining Mass Spectrometry and Nuclear Magnetic Resonance for Metabolomics," *Prog. Nucl. Magn. Reson. Spectrosc.*, vol. 100, pp. 1–16, 2017.

[65] D. S. Wishart *et al.*, "HMDB: The human metabolome database," *Nucleic Acids Res.*, vol. 35, no. SUPPL. 1, pp. 521–526, 2007.

[66] S. C. Gates and C. C. Sweeley, "Quantitative metabolic profiling based on gas chromatography," *Clin. Chem.*, vol. 24, no. 10, pp. 1663–1673, 1978.

[67] B. D. Kell and S. G. Oliver, "The metabolome 18 years on: a concept comes of age," *Metabolomics*, vol. 12, no. 148, 2016.

[68] J. Kruk *et al.*, "NMR Techniques in Metabolomic Studies: A Quick Overview on Examples of Utilization," *Appl. Magn. Reson.*, 2016.

[69] R. Mandal *et al.*, "Multi-platform characterization of the human cerebrospinal fluid metabolome: a comprehensive and quantitative update," *Genome Med.*, vol. 4, no. 38, pp. 1–11, 2012.

[70] S. Ravanbakhsh *et al.*, "Accurate, fully-automated NMR spectral profiling for metabolomics," *PLoS One*, vol. 10, no. 5, pp. 1–15, 2015.

[71] R. M. Maggio, N. L. Calvo, S. E. Vignaduzzo, and T. S. Kaufman, "Pharmaceutical impurities and degradation products: Uses and applications of NMR techniques," *J. Pharm. Biomed. Anal.*, vol. 101, pp. 102–122, 2014.

[72] J. L. Markley *et al.*, "The future of NMR-based metabolomics," *Curr. Opin. Biotechnol.*, vol. 43, pp. 34–40, 2017.

[73] C. P. Slichter, *Principles of Magnetic Resonance*, 2nd ed. Springer, 1978.

[74] I. I. Rabi, S. Millman, K. P, and S. Zacharias, "The Molecular Bean Resonance Method for Measuring Nuclear Magnetic Moments," *Phys. Rev.*, vol. 55, pp. 527–535, 1939.

[75] A. G. Redfield, "On the Theory of Relaxation Processes," *IBM J. Res. Dev.*, vol. 1, no. 1, pp. 19–31, 1957.

[76] C. H. Yoder and C. D. Schaeffer, *Introduction to Multinuclear NMR*. Menlo Park, California: The Benjamin/Cummings Publishing Company Inc., 1987.

[77] P. J. Hore, *Nuclear Magnetic Resonance*. Oxford University Press, Walton Street, Oxford, OX2 6DP, 1995.

[78] T. D. W. Claridge, *High Resolution NMR in Organic Chemistry*. 2016.

[79] W. Kemp, *NMR in Chemistry: A Multinuclear Introduction*. London: Macmillan education Ltd, 1986.

[80] The NMR Lab, "NMR Lab: Pulse Sequences and the Vector Model," 2015. .

[81] Compound Interest, "Analytical Chemistry – A Guide to Proton Nuclear Magnetic Resonance (NMR)," 2015. .

[82] O. Beckonert *et al.*, "Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts," *Nat Protoc*, vol. 2, no. 11, pp. 2692–2703, 2007.

[83] D. Raftery, "Metabolomics-Based Methods for Early Disease Diagnostics: A Review," *Expert Rev Mol Diagn.*, vol. 8, no. 5, pp. 617–633, 2014.

[84] M. Grootveld, *Metabolic Profiling: Disease and Xenobiotics (Issues in Toxicology)*, 1st ed. Royal Society of Chemistry, 2014.

[85] A. J. P. Martin and R. L. M. Synge, "A new form of chromatogram employing two liquid phases," *Biochem. J.*, vol. 35, no. 12, pp. 1358–1368, Dec. 1941.

[86] C. G. Horvath and S. R. Lipsky, "Use of liquid ion exchange chromatography for the separation of organic compounds ," *Nature*, vol. 211, no. 5050. Nature Publishing Group, pp. 748–749, 1966.

[87] W. Prout, *WITH Correction in the Essay on the Relation between the Specific Gravities of Bodies in their Gaseous State and the Weights of their Atoms*, 1st ed. London C Baldwin, 1816.

[88] F. W. Aston, "LXXIV. A positive ray spectrograph," *London, Edinburgh Dublin Philos. Mag. J. Sci.*, vol. 38, no. 228, pp. 707–714, 1919.

[89] A. Dempster, "A New Method of Positive Ray Analysis," *Phys. Rev.*, vol. 11, no. 4, pp. 316–325, 1918.

[90] F. Pullen, "The fascinating history of the development of LC-MS; a personal perspective," *Chromatogr. Today*, pp. 4–6, 2012.

[91] U. Neue, *HPLC Columns: Theory, Technology, and Practice - 1997 - Wiley Analytical Science*. Wiley Analytical Science, 1997.

[92] J. Xia, N. Psychogios, N. Young, and D. S. Wishart, "MetaboAnalyst: A web server

for metabolomic data analysis and interpretation," *Nucleic Acids Res.*, vol. 37, no. SUPPL. 2, pp. 652–660, 2009.

[93] M. Watanabe, M. E. Suliman, A. R. Qureshi, E. Garcia-lopez, and P. Ba, "Consequences of low plasma histidine in chronic kidney disease patients : associations with inflammation , oxidative stress , and," pp. 1860–1866, 2008.

[94] W. L. Buntine, "Bayesian Back-Propagation," 1991.

[95] D. Daniel and K. Thangavel, "Breathomics for gastric cancer classification using back-propagation neural network," *J. Med. Signals Sens.*, vol. 6, no. 3, pp. 172–182, Jul. 2016.

[96] L. Wilkinson and M. Friendly, "The History of the Cluster Heat Map," *Am. Stat.*, vol. 63, no. 2, pp. 179–184, 2009.

[97] R. H. Barton *et al.*, "The influence of EDTA and citrate anticoagulant addition to human plasma on information recovery from NMR-based metabolic profiling studies," *Mol. Biosyst.*, vol. 6, no. 1, pp. 215–224, 2009.

[98] J. K. Nicholson, P. J. D. Foxall, M. Spraul, R. D. Farrant, and J. C. Lindon, "750-MHz H-1 and H-1-C-13 Nmr-Spectroscopy of Human Blood-Plasma," *Anal. Chem.*, vol. 67, no. 2, pp. 793–811, 1995.

[99] D. S. Wishart *et al.*, "HMDB: the Human Metabolome Database," *Nucleic Acids Res.*, vol. 35, no. Database, pp. D521–D526, Jan. 2007.

[100] I. Laíns *et al.*, "Human plasma metabolomics in age-related macular degeneration (AMD) using nuclear magnetic resonance spectroscopy," *PLoS One*, vol. 12, no. 5, p. e0177749, 2017.

[101] F. Dieterle, A. Ross, tz Schlotterbeck, H. Senn, and F. Hoffman-, "Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1 H NMR Metabonomics," 2006.

[102] M. Carayol *et al.*, "Reliability of serum metabolites over a two-year period: A targeted metabolomic approach in fasting and non-fasting samples from EPIC," *PLoS One*, vol. 10, no. 8, pp. 1–10, 2015.

[103] M. Kriat, S. Confort-Gouny, J. Vion-Dury, M. Sciaky, P. Viout, and P. J. Cozzone, "Quantitation of metabolites in human blood serum by proton magnetic resonance spectroscopy. A comparative study of the use of formate and TSP as concentration standards," *NMR Biomed.*, vol. 5, no. 4, pp. 179–184, Jul. 1992.

[104] J. F. Desforges and E. Beutler, "Gaucher's Disease," *N. Engl. J. Med.*, vol. 325, no. 19, pp. 1354–1360, Nov. 1991.

[105] R. J. Wenstrup, M. Roca-Espiau, N. J. Weinreb, and B. Bembi, "Skeletal aspects of Gaucher disease: a review," *Br. J. Radiol.*, vol. 75, no. suppl_1, pp. A2–A12, May 2002.

[106] G. M. Pastores, G. M. Pastores, N. L. Barnett, P. Bathan, and E. H. Kolodny, "A neurological symptom survey of patients with type I Gaucher disease," *J. Inherit. Metab. Dis.*, vol. 26, no. 7, pp. 641–645, 2003.

[107] P. K. Mistry *et al.*, "Effect of oral eliglustat on splenomegaly in patients with Gaucher disease type 1: The ENGAGE randomized clinical trial," *JAMA - J. Am. Med. Assoc.*, vol. 313, no. 7, pp. 695–706, Feb. 2015.

[108] T. M. Cox and J. P. Schofield, *3 Gaucher's disease: clinical features and natural history*, vol. 10, no. 4. 1997.

[109] M. de Fost *et al.*, "Low HDL cholesterol levels in type I Gaucher disease do not lead to an increased risk of cardiovascular disease," *Atherosclerosis*, vol. 204, no. 1, pp. 267–272, May 2009.

[110] L.-K. Tsai, Y.-H. Chien, C.-C. Yang, and W.-L. Hwu, "Myopathy in Gaucher disease," *J. Inherit. Metab. Dis.*, vol. 31, no. S3, pp. 489–491, Dec. 2008.

[111] T. M. Cox and J. P. Scholfield, *Gaucher ' s disease : clinical features and natural history*, vol. 10, no. 4. 1997.

[112] E. A. T. Thévenot, A. Roux, Y. Xu, E. Ezan, and C. Junot, "Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses," 2015.

[113] A. Nicolescu, N. Simionescu, L. Ursu, C. Deleanu, and B. C. Simionescu, "The effect of therapeutic doses of paracetamol and aspirin on the NMR profile of urine at 400 MHZ," *Rev. Roum. Chim.*, vol. 57, no. 7–8, pp. 653–658, 2012.

[114] J. R. Bales, P. J. Sadler, J. K. Nicholson, and J. A. Timbrell, "Urinary excretion of acetaminophen and its metabolites as studied by proton NMR spectroscopy," *Clin. Chem.*, vol. 30, no. 10, pp. 1631–1636, 1984.

[115] † Elaine Holmes *et al.*, "Detection of Urinary Drug Metabolite (Xenometabolome) Signatures in Molecular Epidemiology Studies via Statistical Total Correlation (NMR) Spectroscopy," 2007.

[116] N. Belmatoug *et al.*, "Management and monitoring recommendations for the use of eliglustat in adults with type 1 Gaucher disease in Europe," *Eur. J. Intern. Med.*, vol. 37, pp. 25–32, Jan. 2017.

[117] L. L. Bennett and K. Turcotte, "Eliglustat tartrate for the treatment of adults with type 1 Gaucher disease.," *Drug Des. Devel. Ther.*, vol. 9, pp. 4639–47, 2015.

[118] A. M. Drachnik *et al.*, "Conversion of Pregabalin to 4-Isobutylpyrrolidone-2," *J. Pharm. Sci.*, vol. 106, no. 10, pp. 3095–3102, Oct. 2017.

[119] F. Madrid-Gambin *et al.*, "Impact of chlorogenic acids from coffee on urine metabolome in healthy human subjects," *Food Res. Int.*, vol. 89, pp. 1064–1070, Nov. 2016.

[120] "High Blood Cholesterol - What You Need to Know | National Heart, Lung, and Blood Institute (NHLBI)." [Online]. Available: https://www.nhlbi.nih.gov/health-topics/all-publications-and-resources/high-blood-cholesterol-what-you-need-know. [Accessed: 30-Dec-2019].

[121] T. H. Taddei *et al.*, "High incidence of cholesterol gallstone disease in type 1 Gaucher disease: characterizing the biliary phenotype of type 1 Gaucher disease," *J. Inherit. Metab. Dis.*, vol. 33, no. 3, pp. 291–300, Jun. 2010.

[122] B. A. M. Schutte *et al.*, "The effect of standardized food intake on the association between BMI and 1 H-NMR metabolites," *Sci. Rep.*, vol. 6, no. June, pp. 1–6, 2016.

[123] H. Ginsberg *et al.*, "Reduced plasma concentrations of total, low density lipoprotein

and high density lipoprotein cholesterol in patients with Gaucher type I disease," *Clin. Genet.*, vol. 26, no. 2, pp. 109–116, Apr. 1984.

[124] N. A. Le, J. C. Gibson, A. Rubinstein, G. A. Grabowski, and H. N. Ginsberg, "Abnormalities in lipoprotein metabolism in Gaucher type 1 disease," *Metabolism*, vol. 37, no. 3, pp. 240–245, Mar. 1988.

[125] A. E. Morgan, K. M. Mooney, S. J. Wilkinson, N. A. Pickles, and M. T. Mc Auley, "Cholesterol metabolism: A review of how ageing disrupts the biological mechanisms responsible for its regulation," *Ageing Res. Rev.*, vol. 27, pp. 108–124, May 2016.

[126] M. D. Ashen and R. S. Blumenthal, "Low HDL Cholesterol Levels," *N. Engl. J. Med.*, vol. 353, no. 12, pp. 1252–1260, Sep. 2005.

[127] M. de Fost *et al.*, "Low HDL cholesterol levels in type I Gaucher disease do not lead to an increased risk of cardiovascular disease.," *Atherosclerosis*, vol. 204, no. 1, pp. 267–72, May 2009.

[128] P. K. Mistry *et al.*, "A reappraisal of Gaucher disease-diagnosis and disease management algorithms," *Am. J. Hematol.*, vol. 86, no. 1, pp. 110–115, Jan. 2011.

[129] M. Pocovi *et al.*, "β-glucocerebrosidase gene locus as a link for Gaucher's disease and familial hypo-α-lipoproteinaemia," *Lancet*, vol. 351, no. 9120, pp. 1919–1923, 1998.

[130] A. Cenarro, M. Pocovi, P. Giraldo, A. L. Garcia-Otin, and J. M. Ordovas, "Plasma lipoprotein responses to enzyme-replacement in Gaucher's disease," *Lancet*, vol. 353, no. 9153, pp. 642–643, Feb. 1999.

[131] N. J. Weinreb, D. S. Barbouth, and R. E. Lee, "Causes of death in 184 patients with type 1 Gaucher disease from the United States who were never treated with enzyme replacement therapy," *Blood Cells, Mol. Dis.*, vol. 68, pp. 211–217, Feb. 2018.

[132] A. Safaei *et al.*, "Metabolomic analysis of human cirrhosis, hepatocellular carcinoma, non-alcoholic fatty liver disease and non-alcoholic steatohepatitis diseases.," *Gastroenterol. Hepatol. from bed to bench*, vol. 9, no. 3, pp. 158–73, 2016.

[133] P. K. Mistry, P. Deegan, A. Vellodi, J. A. Cole, M. Yeh, and N. J. Weinreb, "Timing of initiation of enzyme replacement therapy after diagnosis of type 1 Gaucher disease: effect on incidence of avascular necrosis," *Br. J. Haematol.*, vol. 147, no. 4, pp. 561–570, Nov. 2009.

[134] R. Amathieu *et al.*, "Serum 1H-NMR Metabolomic Fingerprints of Acute-On-Chronic Liver Failure in Intensive Care Unit Patients with Alcoholic Cirrhosis," *PLoS One*, vol. 9, no. 2, p. e89230, Feb. 2014.

[135] R. Amathieu *et al.*, "Metabolomic Approach by 1 H NMR Spectroscopy of Serum for the Assessment of Chronic Liver Failure in Patients with Cirrhosis," *J. Proteome Res*, vol. 10, pp. 3239–3245, 2011.

[136] A. McMillan *et al.*, "Metabolic derangements identified through untargeted metabolomics in a cross-sectional study of Nigerian children with severe acute malnutrition," *Metabolomics*, vol. 13, no. 2, p. 13, Feb. 2017.

[137] A. Reddy, S. Lam, and S. Bauer, "Lactic acidosis: Clinical implications and management strategies," *Cleve. Clin. J. Med.*, vol. 82, pp. 615–624, 2015.

[138] W. Bernal, N. Donaldson, D. Wyncoll, and J. Wendon, "Blood lactate as an early predictor of outcome in paracetamol-induced acute liver failure: a cohort study," *Lancet*, vol. 359, no. 9306, pp. 558–563, Feb. 2002.

[139] O. E. Owen *et al.*, "Acetone metabolism during diabetic ketoacidosis," *Diabetes*, vol. 31, no. 3, pp. 242–248, 1982.

[140] H. M. Liebich and J. Wöll, "Volatile substances in blood serum: Profile analysis and quantitative determination," *J. Chromatogr. A*, vol. 142, no. C, pp. 505–516, Nov. 1977.

[141] L. Zhang *et al.*, "Distinguishing pancreatic cancer from chronic pancreatitis and healthy individuals by 1H nuclear magnetic resonance-based metabonomic profiles," *Clin. Biochem.*, vol. 45, no. 13–14, pp. 1064–1069, Sep. 2012.

[142] M. Y. Morgan, A. W. Marshall, J. P. Milsom, and S. Sherlock, "Plasma amino-acid patterns in liver disease.," *Gut*, vol. 23, no. 5, pp. 362–70, May 1982.

[143] L. D. Osellame *et al.*, "Mitochondria and Quality Control Defects in a Mouse Model of Gaucher Disease—Links to Parkinson's Disease," *Cell Metab.*, vol. 17, no. 6, pp. 941–953, Jun. 2013.

[144] I. Rubio-Aliaga *et al.*, "Metabolomics of prolonged fasting in humans reveals new catabolic markers," *Metabolomics*, vol. 7, no. 3, pp. 375–387, Sep. 2011.

[145] E. Kirk, D. N. Reeds, B. N. Finck, M. S. Mayurranjan, and S. Klein, "Dietary fat and carbohydrates differentially alter insulin sensitivity during caloric restriction."

[146] S. H. Zeisel, "Choline: an essential nutrient for humans," *Nutrition*, vol. 16, no. 7–8, pp. 669–671, Jul. 2000.

[147] D. S. Wishart *et al.*, "HMDB 3.0-The Human Metabolome Database in 2013," *Nucleic Acids Res.*, vol. 41, no. D1, 2013.

[148] D. S. Wishart *et al.*, "HMDB 4.0: The human metabolome database for 2018," *Nucleic Acids Res.*, 2018.

[149] J. Tian *et al.*, "The hematinic effect of Colla corii asini (Ejiao) using ^{1}H-NMR metabolomics coupled with correlation analysis in APH-induced anemic rats," *RSC Adv.*, vol. 7, no. 15, pp. 8952–8962, Jan. 2017.

[150] S. S. Ahmed, W. Santosh, S. Kumar, and H. T. T. Christlet, "Metabolic profiling of Parkinson's disease: evidence of biomarker from gene expression analysis and rapid neural network detection.," *J. Biomed. Sci.*, vol. 16, no. 1, p. 63, Jul. 2009.

[151] L. Parnetti *et al.*, "Increased cerebrospinal fluid pyruvate levels in Alzheimer's disease," *Neurosci. Lett.*, vol. 199, no. 3, pp. 231–233, Oct. 1995.

[152] L. Parnetti *et al.*, "INCREASED CSF PYRUVATE LEVELS AS A MARKER OF IMPAIRED ENERGY METABOLISM IN ALZHEIMER'S DISEASE," *J. Am. Geriatr. Soc.*, vol. 43, no. 3, pp. 316–318, Mar. 1995.

[153] T. T. Nielsen and N. S. Sørensen, "Citrate in Plasma and Urine During Total Fasting," *Acta Med. Scand.*, vol. 205, no. 1–6, pp. 303–307, Apr. 2009.

[154] L. C. Costello and R. B. Franklin, "Plasma Citrate Homeostasis: How It Is Regulated; And Its Physiological and Clinical Implications. An Important, But Neglected, Relationship in Medicine.," *HSOA J. Hum. Endocrinol.*, vol. 1, no. 1, 2016.

[155] P. Mikosch, "Gaucher disease and bone," *Best Pract. Res. Clin. Rheumatol.*, vol. 25, no. 5, pp. 665–681, 2011.

[156] J. Prestel *et al.*, "Clinical and molecular characterisation of a Parkinson family with a novel PINK1 mutation," *J. Neurol.*, vol. 255, no. 5, pp. 643–648, May 2008.

[157] S. K. Byeon, J. Y. Lee, J.-S. Lee, and M. H. Moon, "Lipidomic profiling of plasma and urine from patients with Gaucher disease during enzyme replacement therapy by nanoflow liquid chromatography–tandem mass spectrometry," *J. Chromatogr. A*, vol. 1381, pp. 132–139, Feb. 2015.

[158] M. Masada, Y. Kuroda, N. Terumichi, and T. Uno, "Structural Investigation of New Metabolites of Amino-pencillins excreted in Human Urine," *Chem. Pharm. Bull.*, vol. 28, no. 12, pp. 3527–3536, 1980.

[159] A. Awan, J. Barber, R. K. Brennan, and J. A. Parkinson, "Structural Studies on Clarithomycin (6-O-Methlerthromycin A): Assignments of the 1H and 13C NMR Spectra in Organic and Aqueous Solutions." pp. 1241–1246, 1992.

[160] E. J. Saude *et al.*, "Metabolomic profiling of asthma: Diagnostic utility of urine nuclear magnetic resonance spectroscopy," *J. Allergy Clin. Immunol.*, vol. 127, no. 3, pp. 757-764.e6, Mar. 2011.

[161] V. Arjunan, R. Santhanam, M. K. Marchewka, and S. Mohan, "Comprehensive quantum chemical and spectroscopic (FTIR, FT-Raman, 1H, 13C NMR) investigations of O-desmethyltramadol hydrochloride an active metabolite in tramadol – An analgesic drug," *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 122, pp. 315–330, Mar. 2014.

[162] W. N. Wu, L. A. McKown, A. D. Gauthier, W. J. Jones, and R. B. Raffa, "Metabolism of the analgesic drug, tramadol hydrochloride, in rat and dog," *Xenobiotica*, vol. 31, no. 7, pp. 423–441, Jan. 2001.

[163] K. Zaitsu *et al.*, "Metabolic profiling of urine and blood plasma in rat models of drug addiction on the basis of morphine, methamphetamine, and cocaine-induced conditioned place preference," *Anal. Bioanal. Chem.*, vol. 406, no. 5, pp. 1339–1354, Feb. 2014.

[164] J. S. Bajaj *et al.*, "Systems biology analysis of omeprazole therapy in cirrhosis demonstrates significant shifts in gut microbiota composition and function," *Am. J. Physiol. Liver Physiol.*, vol. 307, no. 10, pp. G951–G957, Nov. 2014.

[165] K. Takeuchi *et al.*, "Metabolomic analysis of the effects of omeprazole and famotidine on aspirin-induced gastric injury," *Metabolomics*, vol. 10, no. 5, pp. 995–1004, Oct. 2014.

[166] Z. Chen and Y. Jiang, "Prediction of Response of Collagen-induced Arthritis Rats to Methotrexate: An 1 H-NMR-based Urine Metabolomic Analysis," *J Huazhong Univ Sci Technol [Med Sci]* , vol. 32, no. 3, p. 2012.

[167] S. Y. Um *et al.*, "Pattern Recognition Analysis for the Prediction of Adverse Effects by Nonsteroidal Anti-Inflammatory Drugs Using ^{1}H NMR-Based Metabolomics in Rats," *Anal. Chem.*, vol. 81, no. 12, pp. 4734–4741, Jun. 2009.

[168] J. Jung *et al.*, "1H NMR-based metabolic profiling of naproxen-induced toxicity in rats," *Toxicol. Lett.*, vol. 200, no. 1–2, pp. 1–7, 2011.

[169] C. U. Niemann and N. J. Serkova, "Biochemical mechanisms of nephrotoxicity: application for metabolomics," *Expert Opin. Drug Metab. Toxicol.*, vol. 3, no. 4, pp. 527–544, Aug. 2007.

[170] C. He *et al.*, "1H NMR based pharmacometabolomics analysis of metabolic phenotype on predicting metabolism characteristics of losartan in healthy volunteers," *J. Chromatogr. B*, vol. 1095, pp. 15–23, Sep. 2018.

[171] Y. Ji *et al.*, "Glycine and a Glycine Dehydrogenase (GLDC) SNP as Citalopram/Escitalopram Response Biomarkers in Depression: Pharmacometabolomics-Informed Pharmacogenomics," 2011.

[172] P. Zheng *et al.*, "Identification of sex-specific urinary biomarkers for major depressive disorder by combined application of NMR- and GC–MS-based metabonomics," *Transl. Psychiatry*, vol. 6, no. 11, pp. e955–e955, Nov. 2016.

[173] X. Yu *et al.*, "A urinary metabolomics study of the metabolic dysfunction and the regulation effect of citalopram in the rats exposed to chronic unpredictable mild stress," *RSC Adv.*, vol. 5, pp. 69800–69812, 2015.

[174] P. Bannon, P. Yu, J. M. Cook, L. Roy, and J. P. Villeneuve, "Identification of quinine metabolites in urine after oral dosing in humans," *J. Chromatogr. B Biomed. Appl.*, vol. 715, no. 2, pp. 387–393, 1998.

[175] H. Deguchi *et al.*, "Warfarin Untargeted Metabolomics Study Identifies Novel Procoagulant Ethanolamide Plasma Lipids," *Br J Haematol*, vol. 165, no. 3, pp. 409–412, 2014.

[176] S. Ferrari *et al.*, "Osteoporosis in young adults: pathophysiology, diagnosis, and management," *Osteoporos. Int.*, vol. 23, no. 12, pp. 2735–2748, Dec. 2012.

[177] J. E. Kim *et al.*, "Metabolomics approach to serum biomarker for loperamide-induced constipation in SD rats," *Lab Anim Res*, vol. 30, no. 1, pp. 35–43, 2014.

[178] X. Song, J. Wang, P. Wang, N. Tian, M. Yang, and L. Kong, "1H NMR-based metabolomics approach to evaluate the effect of Xue-Fu-Zhu-Yu decoction on hyperlipidemia rats induced by high-fat diet," *J. Pharm. Biomed. Anal.*, vol. 78–79, pp. 202–210, 2013.

[179] H. ji Yang *et al.*, "An effective assessment of simvastatin-induced toxicity with NMR-based metabonomics approach," *PLoS One*, vol. 6, no. 2, pp. 1–11, 2011.

[180] H. Vanhanen, Y. A. Kesäniemi, and T. A. Miettinen, "Pravastatin lowers serum cholesterol, cholesterol-precursor sterols, fecal steroids, and cholesterol absorption in man," *Metabolism*, vol. 41, no. 6, pp. 588–595, Jun. 1992.

[181] D. Kofink *et al.*, "Statins Effects of Metabolic Profiles Data from the PREVEND IT (Prevention of Renal and Vascular End-stage Disease Intervention Trial)," *Circ. Cardiovasc. Gent.*, 2017.

[182] M. Wallace *et al.*, "Effects of menstrual cycle phase on metabolomic profiles in premenopausal women."

[183] M. Ruoppolo *et al.*, "Serum metabolomic profiles suggest influence of sex and oral contraceptive use," *Am. J. Transl. Res.*, vol. 6, no. 5, p. 614, 2014.

[184] Q. Chan *et al.*, "Metabolic phenotyping for discovery of urinary biomarkers of diet, xenobiotics and blood pressure in the INTERMAP Study: an overview.," *Hypertens.*

Res., vol. 40, no. 4, pp. 336–345, Apr. 2017.

[185] J. Dyerberg, H. O. Bang, and N. Hjørne, "Fatty acid composition of the plasma lipids in Greenland Eskimos," *Am. J. Clin. Nutr.*, vol. 28, no. 9, pp. 958–966, Sep. 1975.

[186] A. Messori, V. Fadda, D. Maratea, and S. Trippoli, "ω-3 Fatty Acid Supplements for Secondary Prevention of Cardiovascular Disease: From 'No Proof of Effectiveness' to 'Proof of No Effectiveness,'" *JAMA Intern. Med.*, vol. 173, no. 15, p. 1466, Aug. 2013.

[187] M. Schmedes *et al.*, "Lean-seafood intake decreases urinary markers of mitochondrial lipid and energy metabolism in healthy subjects: Metabolomics results from a randomized crossover intervention study," *Mol. Nutr. Food Res.*, vol. 60, no. 7, pp. 1661–1672, Jul. 2016.

[188] J. Bird, P. Calder, M. Eggersdorfer, J. K. Bird, P. C. Calder, and M. Eggersdorfer, "The Role of n-3 Long Chain Polyunsaturated Fatty Acids in Cardiovascular Disease Prevention, and Interactions with Statins," *Nutrients*, vol. 10, no. 6, p. 775, Jun. 2018.

[189] R. A. Jacob *et al.*, "Moderate Folate Depletion Increases Plasma Homocysteine and Decreases Lymphocyte DNA Methylation in Postmenopausal Women," *J. Nutr.*, vol. 128, no. 7, pp. 1204–1212, Jul. 1998.

[190] X. Cui *et al.*, "Analysis of MTHFR, CBS, Glutathione, Taurine, and Hydrogen Sulfide Levels in Retinas of Hyperhomocysteinemic Mice," *Retin. Cell Biol.*, vol. 58, no. 4, pp. 1954–1963, 2017.

[191] J. Godzien *et al.*, "Metabolomic Approach with LC-QTOF to Study the Effect of a Nutraceutical Treatment on Urine of Diabetic Rats."

[192] C. M. Walsh, L. Brennan, J. P. G. Malthouse, H. Roche, and M. Gibney, "Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans," *Am. J. Clin. Nutr.*, vol. 84, no. 3, pp. 531–539, 2006.

[193] M. Gibney, C. M. Walsh, L. Brennan, H. Roche, B. German, and B. V Ommen, "Metabolomics in human nutrition: opportunities and challenges," *Am. J. Clin. Nutr.*, vol. 82, no. 3, pp. 498–503, 2005.

[194] R. Brauer, A. Benedikt, L. • Georg, M. Fiedler, J. Thiery, and U. Ceglarek, "Preanalytical standardization of amino acid and acylcarnitine metabolite profiling in human blood using tandem mass spectrometry."

[195] Z. Reiner *et al.*, "ESC/EAS Guidelines for the management of dyslipidaemias: the Task Force for the management of dyslipidaemias of the European Society of Cardiology (ESC) and the European Atherosclerosis Society (EAS).," *Eur. Heart J.*, vol. 32, no. 14, pp. 1769–818, 2011.

[196] P. Yin, R. Lehmann, and G. Xu, "Effects of pre-analytical processes on blood samples used in metabolomics studies," *Anal. Bioanal. Chem.*, vol. 407, no. 17, pp. 4879–4892, 2015.

[197] M. Carayol *et al.*, "Reliability of Serum Metabolites over a Two-Year Period: A Targeted Metabolomic Approach in Fasting and Non-Fasting Samples from EPIC," *PLoS One*, vol. 10, no. 8, p. e0135437, Aug. 2015.

[198] S. E. Bianconi *et al.*, "Evaluation of age of death in Niemann-Pick disease, type C: Utility of disease support group websites to understand natural history," *Mol. Genet. Metab.*, vol. 126, no. 4, pp. 466–469, Apr. 2019.

[199] H. S. Kruth *et al.*, "Type C Niemann-Pick disease. Abnormal metabolism of low density lipoprotein in homozygous and heterozygous fibroblasts," *J. Biol. Chem.*, vol. 261, no. 35, pp. 16769–16774, 1986.

[200] M. C. Patterson, D. Vecchio, H. Prady, L. Abel, and J. E. Wraith, "Miglustat for treatment of Niemann-Pick C disease: a randomised controlled study," *Lancet Neurol.*, vol. 6, no. 9, pp. 765–772, Sep. 2007.

[201] F. Probert *et al.*, "Urinary excretion and metabolism of miglustat and valproate in patients with Niemann-Pick type C1 disease: One- and two-dimensional solution-state 1H NMR studies," *J. Pharm. Biomed. Anal.*, vol. 117, pp. 276–288, 2016.

[202] C. M. Ramirez *et al.*, "Weekly Cyclodextrin Administration Normalizes Cholesterol Metabolism in Nearly Every Organ of the Niemann-Pick Type C1 Mouse and Markedly Prolongs Life," *Pediatr. Res.*, vol. 68, no. 4, pp. 309–315, Oct. 2010.

[203] A. I. Rosenbaum, G. Zhang, J. D. Warren, and F. R. Maxfield, "Endocytosis of beta-cyclodextrins is responsible for cholesterol reduction in Niemann-Pick type C mutant cells.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 12, pp. 5477–82, Mar. 2010.

[204] S. Ward, P. O'Donnell, S. Fernandez, and C. H. Vite, "2-Hydroxypropyl-β-Cyclodextrin Raises Hearing Threshold in Normal Cats and in Cats With Niemann-Pick Type C Disease," *Pediatr. Res.*, vol. 68, no. 1, pp. 52–56, Jul. 2010.

[205] E. Berry-Kravis *et al.*, "Long-Term Treatment of Niemann-Pick Type C1 Disease With Intrathecal 2-Hydroxypropyl-β-Cyclodextrin.," *Pediatr. Neurol.*, vol. 80, pp. 24–34, Mar. 2018.

[206] T. J. Maarup *et al.*, "Intrathecal 2-hydroxypropyl-beta-cyclodextrin in a single patient with Niemann-Pick C1," *Mol. Genet. Metab.*, vol. 116, no. 1–2, pp. 75–79, Sep. 2015.

[207] C. H. Vite *et al.*, "Intracisternal cyclodextrin prevents cerebellar dysfunction and Purkinje cell death in feline Niemann-Pick type C1 disease," *Sci. Transl. Med.*, vol. 7, no. 276, p. 276ra26, 2015.

[208] N. Mattsson *et al.*, "Amyloid-β metabolism in Niemann-Pick C disease models and patients," *Metab. Brain Dis.*, vol. 27, no. 4, pp. 573–585, Dec. 2012.

[209] S. Rezzi, F. A. Vera, F. P. J. Martin, S. Wang, D. Lawler, and S. Kochhar, "Automated SPE-RP-HPLC fractionation of biofluids combined to off-line NMR spectroscopy for biomarker identification in metabonomics," *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.*, vol. 871, no. 2, pp. 271–278, 2008.

[210] S.-M. Rivera-Velez and F. Villarino, Nicolas, "Feline urine metabolomic signature: characterization of low-molecularweight substances in urine from domestic cats," *J. Feline Med.* , vol. 20, no. 2, pp. 155–163, 2018.

[211] J. Chong *et al.*, "MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis," *Nucleic Acids Res.*, vol. 46, no. W1, pp. W486–W494, Jul. 2018.

[212] A.-H. Emwas *et al.*, "Recommended strategies for spectral processing and post-

processing of 1D 1H-NMR data of biofluids with a particular focus on urine.," *Metabolomics*, vol. 14, no. 3, p. 31, 2018.

[213] K. L. Somers *et al.*, "Effects of dietary cholesterol restriction in a feline model of Niemann–Pick type C disease," *J. Inherit. Metab. Dis.*, vol. 24, no. 4, pp. 427–436, Jul. 2001.

[214] W. H. Hendriks, M. F. Tarttelin, and P. J. Moughan, "Twenty-four hour feline excretion patterns in entire and castrated cats," *Physiol. Behav.*, vol. 58, no. 3, pp. 467–469, Sep. 1995.

[215] W. H. Hendriks, K. J. Rutherfurd-Markwick, K. Weidgraaf, C. Ugarte, and Q. R. Rogers, "Testosterone increases urinary free felinine, N-acetylfelinine and methylbutanolglutathione excretion in cats (Felis catus)," *J. Anim. Physiol. Anim. Nutr. (Berl).*, vol. 92, no. 1, pp. 53–62, Feb. 2008.

[216] F. Dieterle, G. Schlotterbeck, M. Binder, A. Ross, L. Suter, and H. Senn, "Application of Metabonomics in a Comparative Profiling Study Reveals N-Acetylfelinine Excretion as a Biomarker for Inhibition of the Farnesyl Pathway by Bisphosphonates."

[217] C. F. Roff *et al.*, "The murine Niemann-Pick type C lesion affects testosterone production," *Endocrinology*, vol. 133, no. 6, pp. 2913–2923, 1993.

[218] S. Rezzi, F. Arce Vera, F. O.-P. J. Martin, S. Wang, D. Lawler, and S. Kochhar, "Automated SPE-RP-HPLC fractionation of biofluids combined to off-line NMR spectroscopy for biomarker identification in metabonomics," *J. Chromatogr. B*, vol. 871, pp. 271–278, 2008.

[219] M. Miyazaki *et al.*, "A Major Urinary Protein of the Domestic Cat Regulates the Production of Felinine, a Putative Pheromone Precursor," *Chem. Biol.*, vol. 13, no. 10, pp. 1071–1079, Oct. 2006.

[220] N. G. Psihogios, I. F. Gazi, M. S. Elisaf, K. I. Seferiadis, and E. T. Bairaktari, "Gender-related and age-related urinalysis of healthy subjects by NMR-based metabonomics."

[221] R. E. Williams, E. M. Lenz, J. S. Lowden, M. Rantalainen{, and I. D. Wilson, "The metabonomics of aging and development in the rat: an investigation into the effect of age on the profile of endogenous metabolites in the urine of male rats using 1 H NMR and HPLC-TOF MS."

[222] E. M. Lenz *et al.*, "Metabonomics, dietary influences and cultural differences: A 1H NMR-based study of urine samples obtained from healthy British and Swedish subjects," *J. Pharm. Biomed. Anal.*, vol. 36, no. 4, pp. 841–849, 2004.

[223] M. Janeiro *et al.*, "Implication of Trimethylamine N-Oxide (TMAO) in Disease: Potential Biomarker or New Therapeutic Target," *Nutrients*, vol. 10, no. 10, p. 1398, Oct. 2018.

[224] R. A. Koeth *et al.*, "γ-Butyrobetaine is a proatherogenic intermediate in gut microbial metabolism of L-carnitine to TMAO.," *Cell Metab.*, vol. 20, no. 5, pp. 799–812, Nov. 2014.

[225] R. A. Koeth *et al.*, "Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis.," *Nat. Med.*, vol. 19, no. 5, pp. 576–85, May 2013.

[226] Z. Wang *et al.*, "Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease," *Nature*, vol. 472, no. 7341, pp. 57–63, Apr. 2011.

[227] J. Waterfield C, A. Turton J, D. Scales M, and A. Timbrell, J, "Taurine, a possible urinary marker of liver damage: a study of taurine excretion in carbon tetrachloride-treated rats.," *Arch Toxicol.*, vol. 65, no. 7, pp. 548–555, 1991.

[228] E. Holmes, J. V Li, T. Athanasiou, H. Ashrafian, and J. K. Nicholson, "Understanding the role of gut microbiomeâ€"host metabolic signal disruption in health and disease," *Trends Microbiol.*, vol. 19, pp. 349–359, 2011.

[229] A. Tamura and N. Yui, "Polyrotaxane-based systemic delivery of β-cyclodextrins for potentiating therapeutic efficacy in a mouse model of Niemann-Pick type C disease," *J. Control. Release*, vol. 269, pp. 148–158, Jan. 2018.

[230] H. R. Williams *et al.*, "Differences in gut microbial metabolism are responsible for reduced hippurate synthesis in Crohn's disease," 2010.

[231] J. Delaney, W. A. Neville, A. Swain, A. Miles, M. S. Leonard, and C. J. Waterfield, "Phenylacetylglycine, a putative biomarker of phospholipidosis: Its origins and relevance to phospholipid accumulation using amiodarone treated rats as a model," *Biomarkers*, vol. 9, no. 3, pp. 271–290, Jan. 2004.

[232] L. Doessegger *et al.*, "Increased levels of urinary phenylacetylglycine associated with mitochondrial toxicity in a model of drug-induced phospholipidosis," *Ther. Adv. Drug Saf.*, vol. 4, no. 3, pp. 101–114, Jun. 2013.

[233] R. G. Kristiansen *et al.*, "L-Ornithine phenylacetate reduces ammonia in pigs with acute liver failure through phenylacetylglycine formation: a novel ammonia-lowering pathway," *Am. J. Physiol. Liver Physiol.*, vol. 307, no. 10, pp. G1024–G1031, Nov. 2014.

[234] B. E. Kennedy, C. T. Madreiter, N. Vishnu, R. Malli, W. F. Graier, and B. Karten, "Adaptations of energy metabolism associated with increased levels of mitochondrial cholesterol in Niemann-Pick type C1-deficient cells.," *J. Biol. Chem.*, vol. 289, no. 23, pp. 16278–89, Jun. 2014.

[235] S. Gould and R. C. Scott, "2-Hydroxypropyl-??-cyclodextrin (HP-??-CD): A toxicology review," *Food Chem. Toxicol.*, vol. 43, no. 10, pp. 1451–1459, 2005.

[236] N. Brunetti-Pierri and F. Scaglia, "GM1 gangliosidosis: review of clinical, molecular, and therapeutic aspects," *Mol. Genet. Metab.*, vol. 94, no. 4, pp. 391–396, 2008.

[237] U. F. H. Engelke *et al.*, "N-acetylated metabolites in urine: proton nuclear magnetic resonance spectroscopic study on patients with inborn errors of metabolism.," *Clin. Chem.*, vol. 50, no. 1, pp. 58–66, Jan. 2004.

[238] J. Pinto *et al.*, "Human plasma stability during handling and storage: impact on NMR metabolomics," *Analyst*, vol. 139, no. 5, pp. 1168–1177, Feb. 2014.

[239] E. M. Komoroski, R. a Komoroski, J. L. Valentine, J. M. Pearce, and G. L. Kearns, "The use of nuclear magnetic resonance spectroscopy in the detection of drug intoxication.," *J. Anal. Toxicol.*, vol. 24, no. 3, pp. 180–187, 2000.

[240] J. E. Hokanson and M. A. Austin, "Plasma Triglyceride Level is a Risk Factor for Cardiovascular Disease Independent of High-Density Lipoprotein Cholesterol Level: A Metaanalysis of Population-Based Prospective Studies," *Eur. J. Cardiovasc.*

Prev. Rehabil., vol. 3, no. 2, pp. 213–219, Apr. 1996.

[241] U. F. H. Engelke, "N-Acetylated Metabolites in Urine: Proton Nuclear Magnetic Resonance Spectroscopic Study on Patients with Inborn Errors of Metabolism," *Clin. Chem.*, vol. 50, no. 1, pp. 58–66, 2004.

[242] L. iVONA Stefan, A. Nicolescu, S. Popa, M. Mota, E. Kovacs, and C. Deleanu, "1H NMR Urine Metabolic Profiling in Type 1 Diabetes Mellitus," *Rev. Roum. Chim.*, vol. 55, no. 11–12, pp. 1033–1037, 2010.

[243] M. Fulop, J. Bock, J. Ben-Ezra, M. Antony, J. Danzig, and J. S. Gage, "Plasma lactate and 3-hydroxybutyrate levels in patients with acute ethanol intoxication," *Am. J. Med.*, vol. 80, no. 2, pp. 191–194, 1986.

[244] A. Zeharia *et al.*, "Acute Infantile Liver Failure Due to Mutations in the TRMU Gene," *Am. J. Hum. Genet.*, vol. 85, no. 3, pp. 401–407, Sep. 2009.

[245] Y. Qiu *et al.*, "Serum Metabolite Profiling of Human Colorectal Cancer Using GC-TOFMS and UPLC-QTOFMS."

[246] Z. Hu *et al.*, "1H NMR-based metabonomic analysis of brain in rats of morphine dependence and withdrawal intervention," *Behav. Brain Res.*, vol. 231, no. 1, pp. 11–19, May 2012.

[247] S. Quasthoff, C. Möckel, W. Zieglgänsberger, and W. Schreibmayer, "Tolperisone: A Typical Representative of a Class of Centrally Acting Muscle Relaxants with Less Sedative Side Effects," *CNS Neurosci. Ther.*, vol. 14, no. 2, pp. 107–119, Jun. 2008.

[248] P. R. West *et al.*, "Metabolomics as a Tool for Discovery of Biomarkers of Autism Spectrum Disorder in the Blood Plasma of Children," *PLoS One*, vol. 9, no. 11, p. e112445, Nov. 2014.

[249] L. Schultz *et al.*, "Toxicology in Vitro Evaluation of drug-induced neurotoxicity based on metabolomics , proteomics and electrical activity measurements in complementary CNS in vitro models," *Toxicol. Vitr.*, vol. 30, no. 1, pp. 138–165, 2015.

[250] X. Wang *et al.*, "Extending the detection window of diazepam by directly analyzing its glucuronide metabolites in human urine using liquid chromatography–tandem mass spectrometry," *J. Chromatogr. A*, vol. 1268, pp. 29–34, Dec. 2012.

[251] J. F. G. Iii *et al.*, "Metabolomic Analysis Reveals Extended Metabolic Consequences of Marginal Vitamin B-6 Deficiency in Healthy Human Subjects," vol. 8, no. 6, 2013.

[252] C. S. Moore *et al.*, "Oily fish reduces plasma triacylglycerols: a primary prevention study in overweight men and women," *Nutrition*, vol. 22, no. 10, pp. 1012–1024, Oct. 2006.

[253] Y.-Y. Yeh and S.-M. Yeh, "Garlic reduces plasma lipids by inhibiting hepatic cholesterol and triacylglycerol synthesis," *Lipids*, vol. 29, no. 3, pp. 189–193, Mar. 1994.

[254] H. Gu *et al.*, "H NMR metabolomics study of age profiling in children."

[255] I. Montoliu *et al.*, "Serum profiling of healthy aging identifies phospho- and sphingolipid species as markers of human longevity.," *Aging (Albany. NY).*, vol. 6, no. 1, pp. 9–25, Jan. 2014.

[256] J. Tiao, "The effect of age on serum creatinine levels in an aging population: relevance to vascular surgery," *Cardiovasc. Surg.*, vol. 10, no. 5, pp. 445–451, Oct. 2002.

[257] J. Xia, D. I. Broadhurst, M. Wilson, and D. S. Wishart, "Translational biomarker discovery in clinical metabolomics: An introductory tutorial," *Metabolomics*, vol. 9, no. 2, pp. 280–299, 2013.

[258] M. J. Parnham, V. E. Haber, E. J. Giamarellos-Bourboulis, G. Perletti, G. M. Verleden, and R. Vos, "Azithromycin: Mechanisms of action and their relevance for clinical applications," *Pharmacol. Ther.*, vol. 143, no. 2, pp. 225–245, 2014.

[259] Veterinary Medicines Directorate DEFRA, "Product Information Database - Clavaseptin 50mg Palatable Tablets for Dogs and Cats," 2018. [Online]. Available: https://www.vmd.defra.gov.uk/productinformationdatabase/Default.aspx. [Accessed: 23-May-2019].

[260] D. B. Hirata *et al.*, "Precipitation of clavulanic acid from fermentation broth with potassium 2-ethyl hexanoate salt," *Sep. Purif. Technol.*, vol. 66, no. 3, pp. 598–605, May 2009.

[261] A. Baranova, B. Tzu Huang, A. Kocak, E. Champeil John Jay, E. Champeil, and J. Jay, "Quantitation of Amoxicillin in Urine by Nuclear Magnetic Resonance. Application to Five Cases," *J. Clin. Anal. Med.*

[262] T. REYNS *et al.*, "Influence of administration route on the biotransformation of amoxicillin in the pig," *J. Vet. Pharmacol. Ther.*, vol. 32, no. 3, pp. 241–248, Jun. 2009.

[263] G. C. Bolton, G. D. Allen, C. W. Filer, and D. J. Jeffery, "Absorption, metabolism and excretion studies on clavulanic acid in the rat and dog," *Xenobiotica fate foreign Compd. Biol. Syst.*, vol. 14, no. 6, pp. 483–490, 1984.

[264] G. C. Bolton, G. D. Allen, B. E. Davies, C. W. Filer, and D. J. Jeffery, "The disposition of clavulanic acid in man," *Xenobiotica.*, vol. 16, no. 9, pp. 853–863, 1985.

[265] Á. Jerzsele and G. Nagy, "THE STABILITY OF AMOXICILLIN TRIHYDRATE AND POTASSIUM CLAVULANATE COMBINATION IN AQUEOUS SOLUTIONS," *Acta Vet. Hung.*, vol. 57, no. 4, pp. 485–493, 2009.

[266] M. J. Jessa, D. A. Barrett, P. N. Shaw, and R. C. Spiller, "Rapid and selective high-performance liquid chromatographic method for the determination of metronidazole and its active metabolite in human plasma, saliva and gastric juice," *J. Chromatogr. B*, vol. 677, no. 2, pp. 374–379, 1996.

[267] S. A. Dingsdag and N. Hunter, "Metronidazole: an update on metabolism, structure–cytotoxicity and resistance mechanisms," *J. Antimicrob. Chemother.*, vol. 73, no. 2, pp. 265–279, Feb. 2018.

Appendices

Appendix 1 – CPMG *versus* NOESY PRESAT GD1 Samples

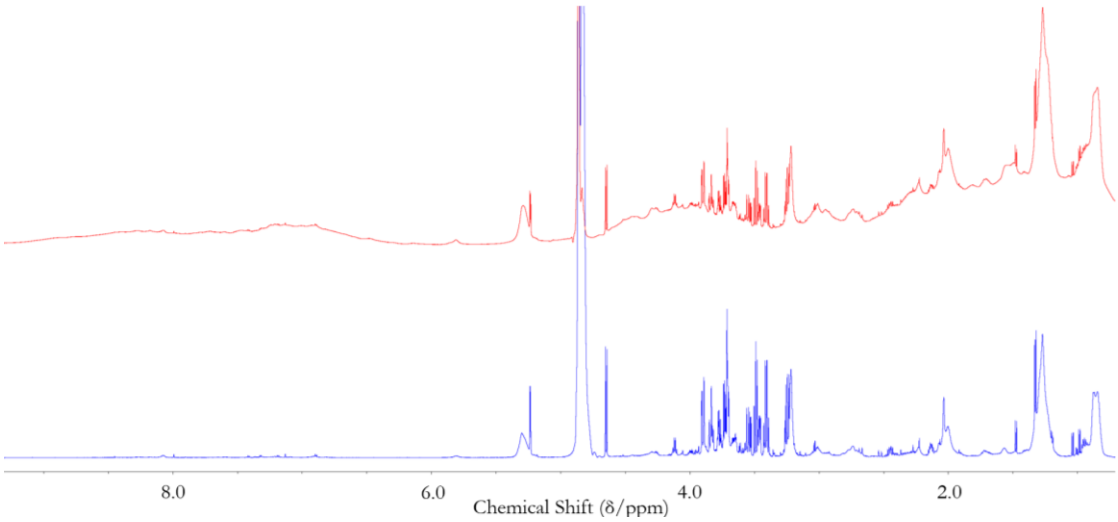


Figure 1. Top: Typical NOESY PRESAT spectrum (red) with clear distortions to the baseline as a result of the protein in the plasma sample. Bottom: CPMG pulse sequence of the same samples (blue) with a clear straight baseline.

Appendix 2 – Outlier identified by PCA in GD1 Plasma

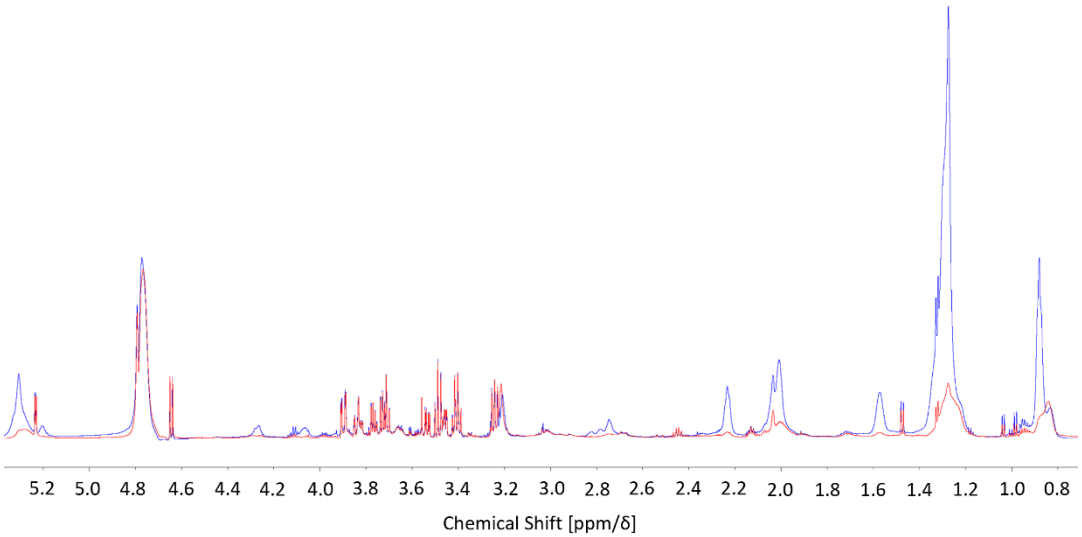


Figure 2. Sample 70P (blue) against a typical spectrum from the control group (red).

Clear outlier in the control group with high lipid resonances observable at δ = 0.88 ppm, δ = 1.33 ppm, δ = 1.55 ppm, δ = 2.01 ppm, δ = 2.22 ppm and δ = 5.25 ppm. This outlier was removed from further statistical analysis.

Appendix 3 – Univariate Data Analysis in GD1 Plasma

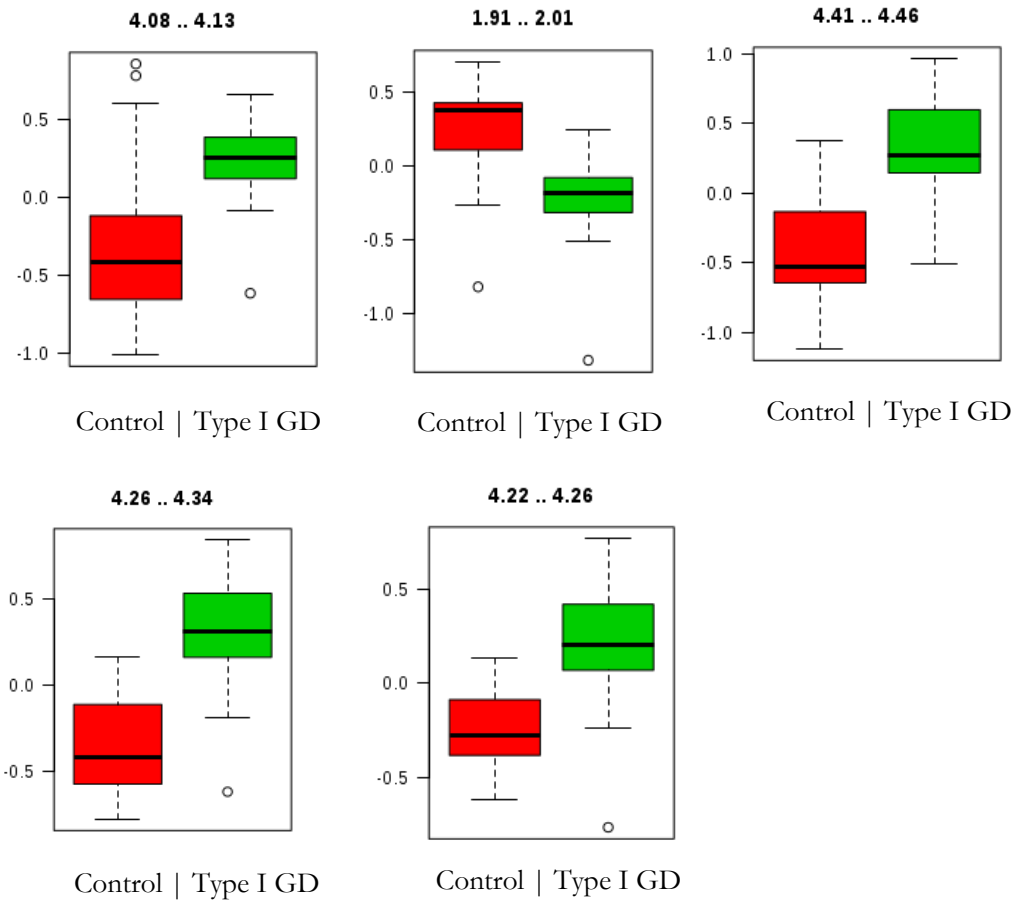


Figure 3a. Box plots of significant buckets in MVA analysis in PCA plots. Buckets include 4.26-4.34, Glyceryl of lipids (*broad*), 4.22-4.26, Threonine-CH /glyceryl of lipids (*m*) 1.91-2.01, $CH_2CH_2CH=$ (*broad*), 4.41-4.46 Unassigned broad singlet (*s*) and 4.08-4.13 Lactate-CH (*q*)/ 3-hydroxybuyrate-CH (*m*).

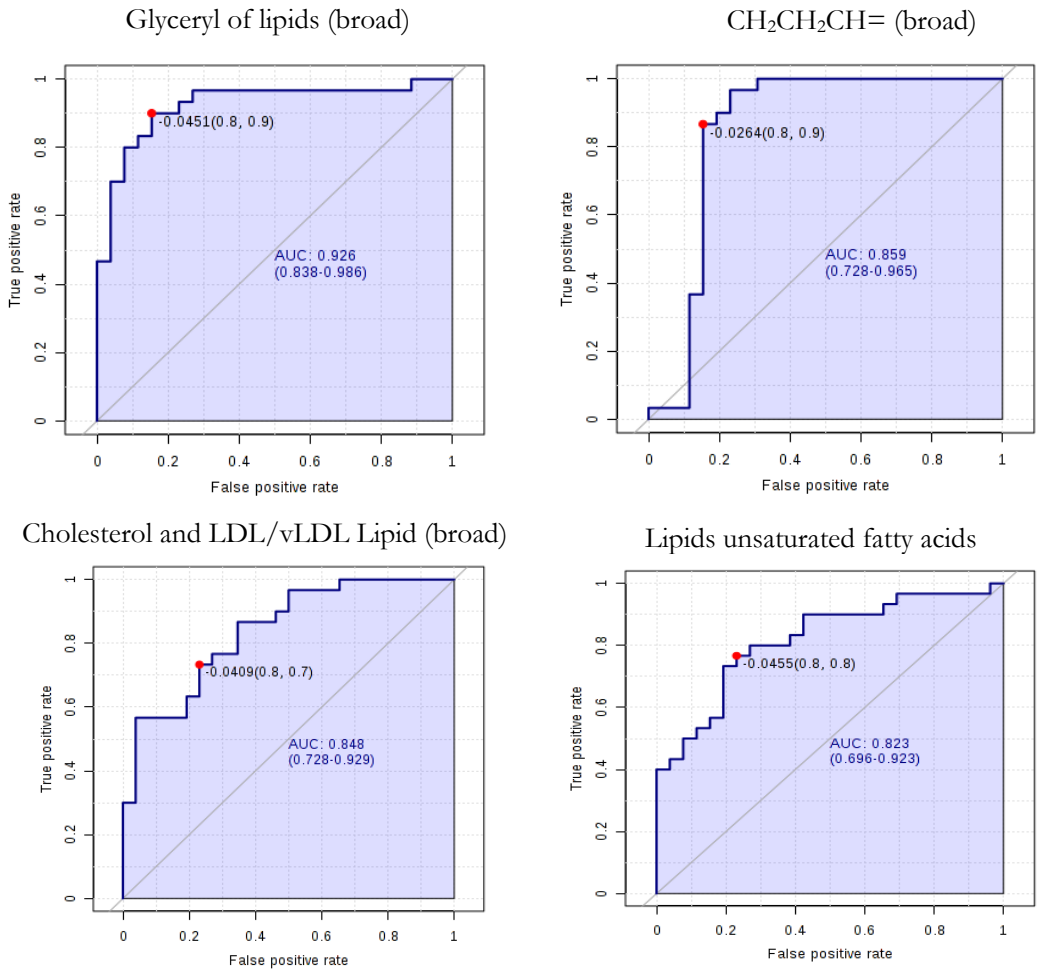


Figure 3b. Univariate ROC Curve analysis for significant metabolites showing false positive rate on the x-axis *vs* true positive rate on the y-axis. These metabolites are comparing controls *vs* GD.

Appendix 4 – Signals Attributable to Drug Metabolites in GD1 Urine

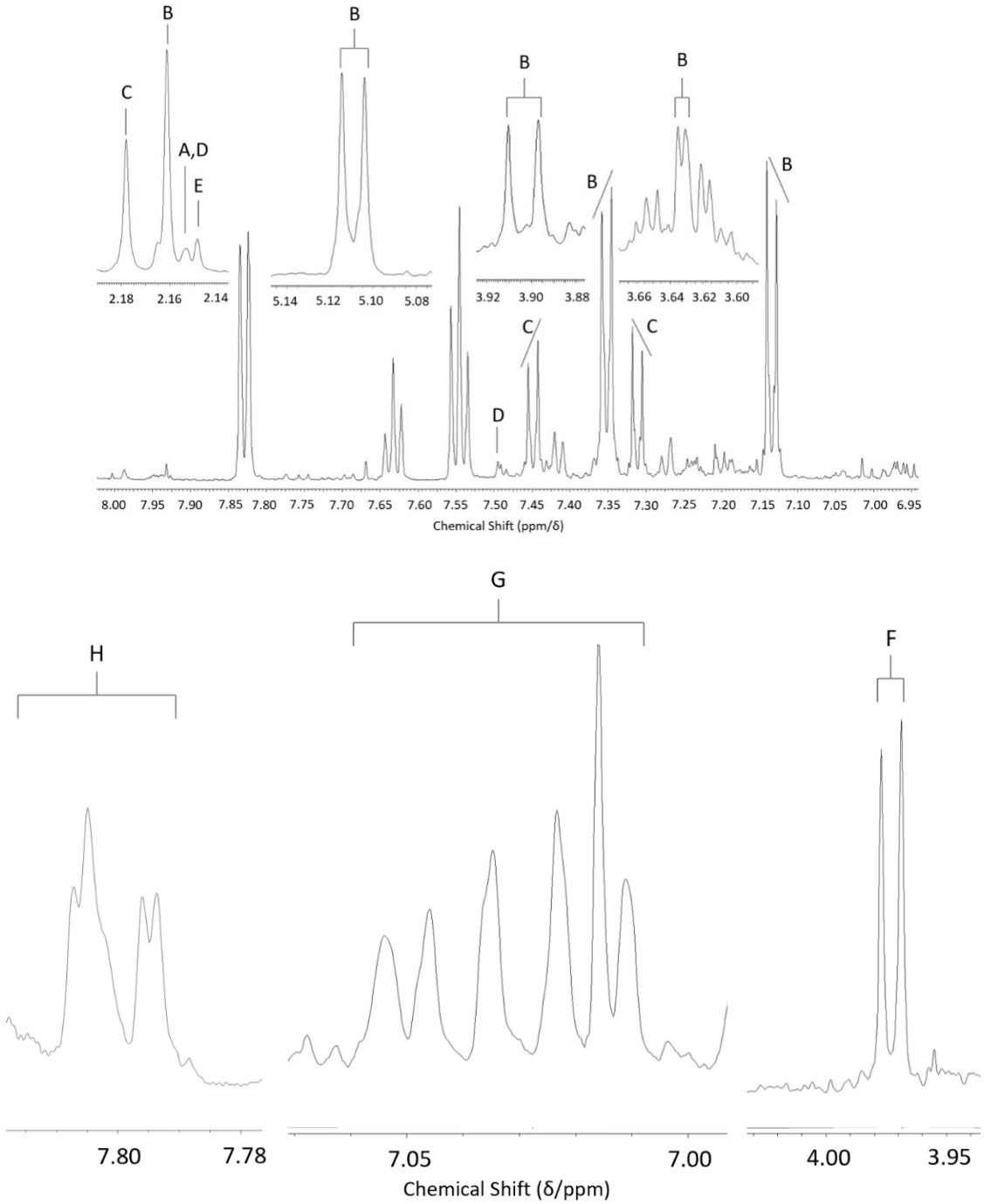


Figure 4. Top: Presence of key acetaminophen and corresponding resonances in human urine from GD patient labelled [A] acetaminophen [B] glucuronide derivative [C] sulphate derivative [D] L-cysteinyl derivative [E] N-acetyl-L cysteinyl derivative Bottom: Assignments in salicyluric acid assigned in human urine in patients with GD [F] salicyluric acid -CH_2 [G] salicyluric acid -CH [H] salicyluric acid -CH.

Appendix 5 – Further Multivariate Data Analysis in GD1 Urine

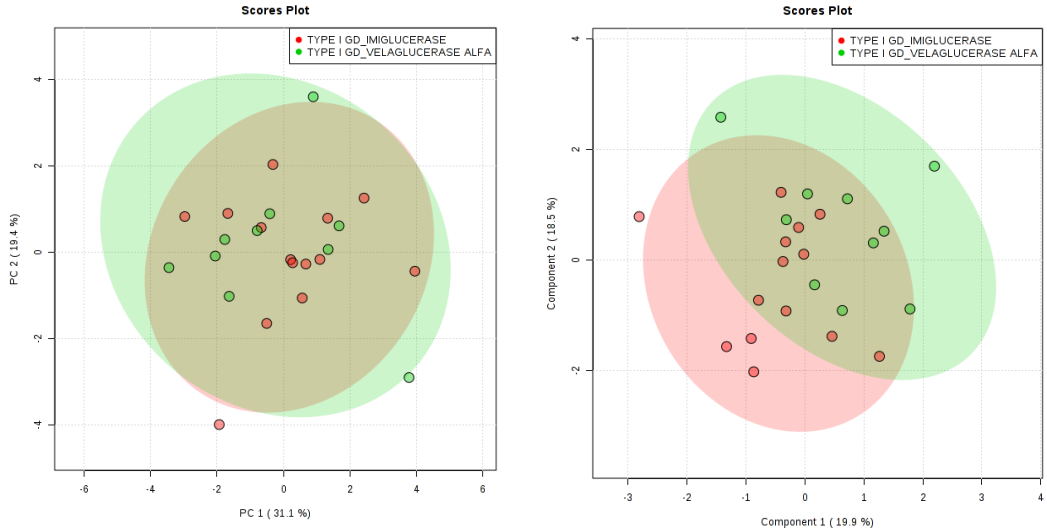


Figure 5a. Left: PCA plot of GD urine on comparison of two treatment types in urinary profiles, with imiglucerase (red) and velaglucerase alfa (green). PC1 *vs* PC2 shown representing 31.1% and 19.4% of the dataset variation respectively. **Right:** The PLS-DA plot shows PC1 *vs* PC2 which represent 19.9% and 18.5% of the dataset variation respectively. Q2, R2 and Accuracy are all less than >0.5, 2000 permutations, p>0.05. The plot shows 95% confidence ellipses.

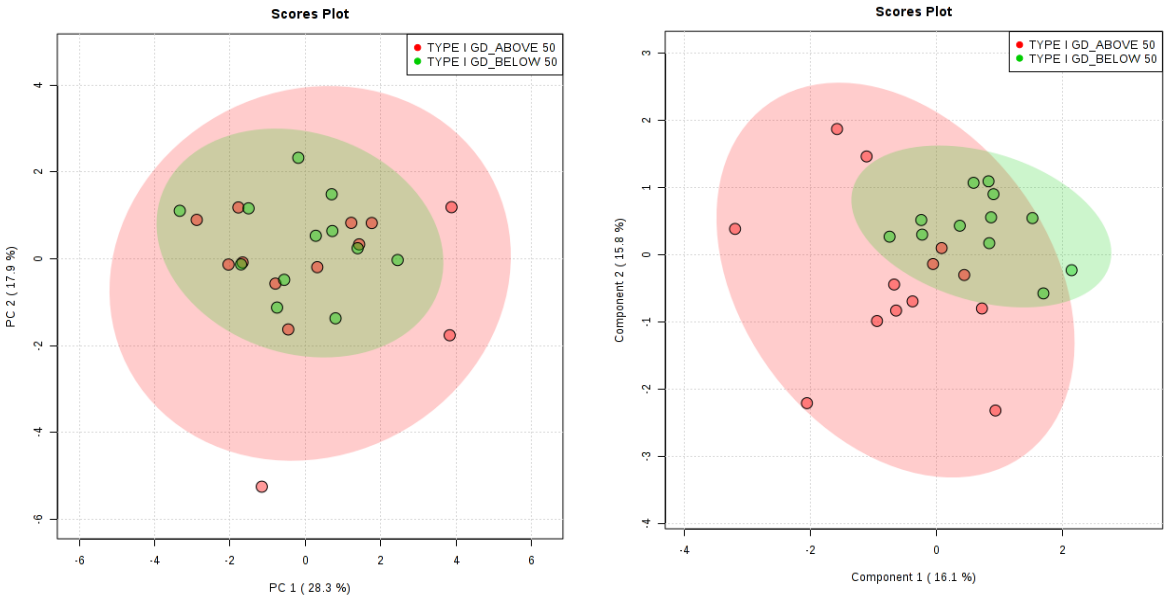


Figure 5b. Left: PCA plot of GD urinary loadings with participants above 50 (red) and below 50 (green). PC1 and PC2 show 28.3% and 17.9% of the dataset variance respectively. **Right:** PLS-DA plot of GD urinary loadings with participants above 50 (red) and below 50 (green). PC1 and PC2 show 16.1% and 15.8% dataset variance respectively. Q^2, R^2 and accuracy are all less than >0.5, 2000 permutations, p>0.05. 95% confidence ellipses are shown for both statistical diagrams.

Appendix 6 – Significant Metabolite in Univariate testing in GD1 Urine

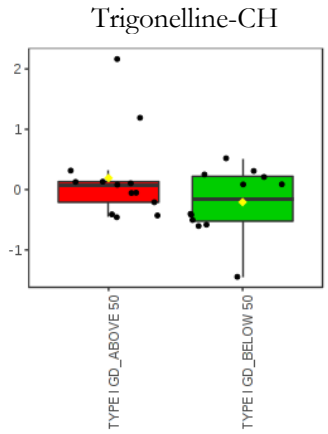


Figure 6. fold change analysis of the signal at δ = 9.10-9.15 ppm attributable to trigonelline

Appendix 7 – Excluded Bucket Regions for Multivariate Analysis in Feline Urine

Anomalous signals

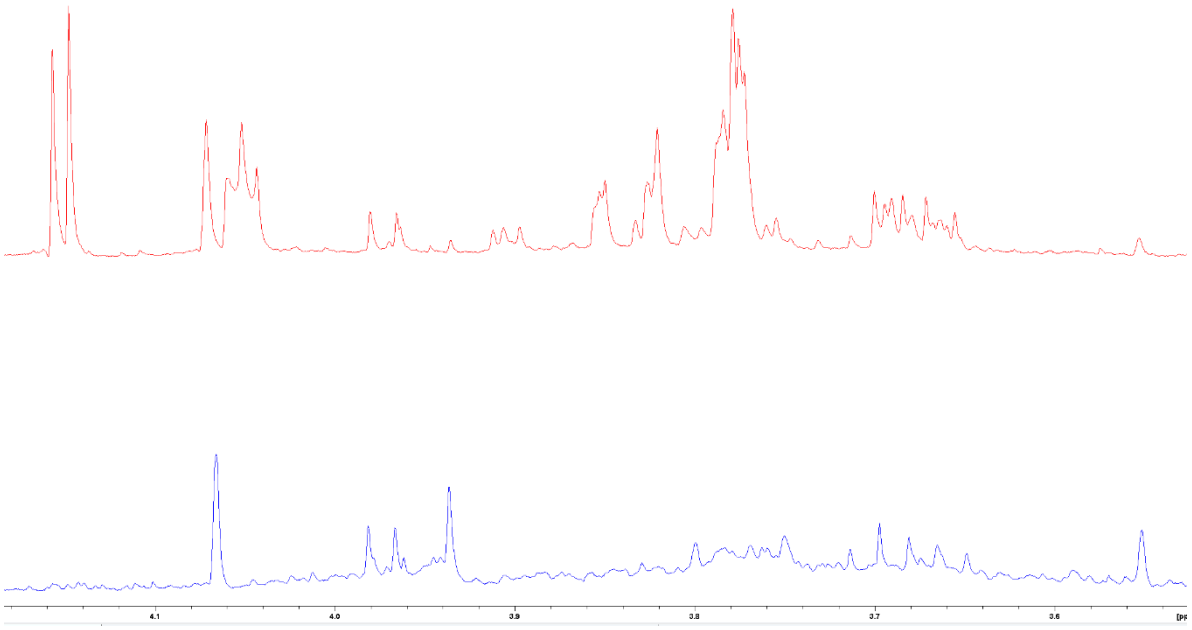


Figure 7.0. Resonances from two untreated NPC felines at 8 weeks of age. Resonances appear in the red spectrum which are not present in the blue spectrum at δ = 3.69 ppm (*m*), 3.77 ppm (*m*), 4.05 ppm (*t*), 4.14 and 4.11 ppm (*d*).

The above spectra show the presence of some anomalous signals in the feline urine samples. These signals are observed in urine from untreated control felines, HPβCD treated felines, treated felines, untreated NPC felines and antibiotic treated felines.

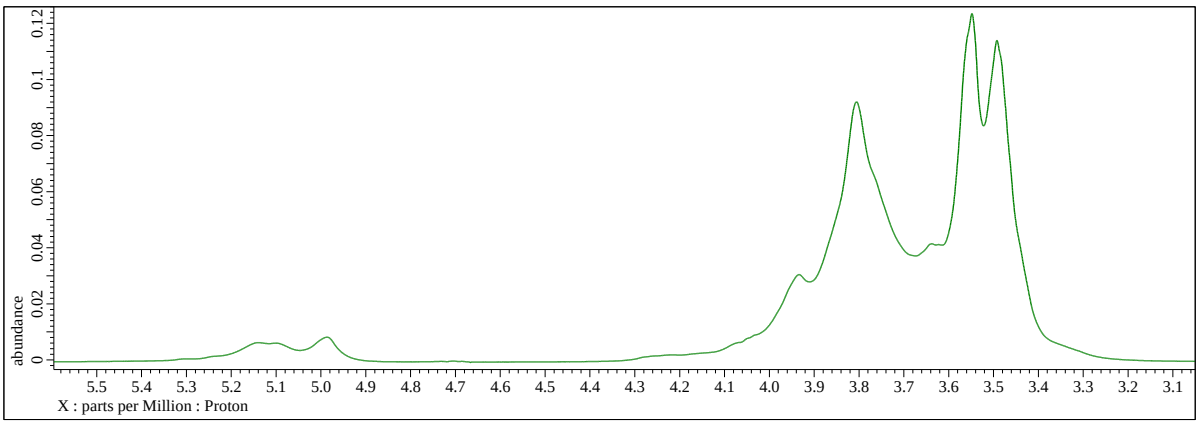


Figure 7.1. Resonances from HPβCD only. Purchased from Sigma Aldrich.

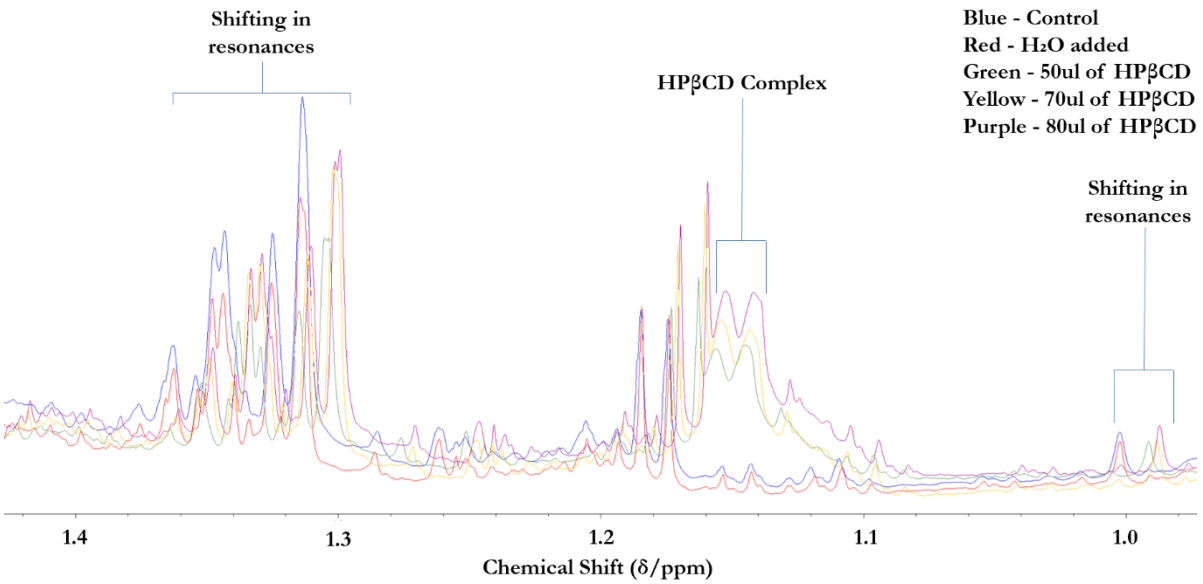


Figure 7.2. HPβCD spiked into control feline urine. Complex formed at δ = 1.4 ppm. No complexation was observed in feline urine from controls and treated felines in the study. Shifting resonances are observable as a result of the addition of HPβCD presence due to complexation of HPβCD with the metabolites. No shifting is observed in this investigation using 252 feline urine samples of which 3 different treatment strategies are employed. The higher the concentration of HPβCD the more signals appear to shift to the right of the spectrum.

Appendix 8 – Drug Metabolism in Felines

Drug treated feline spectra

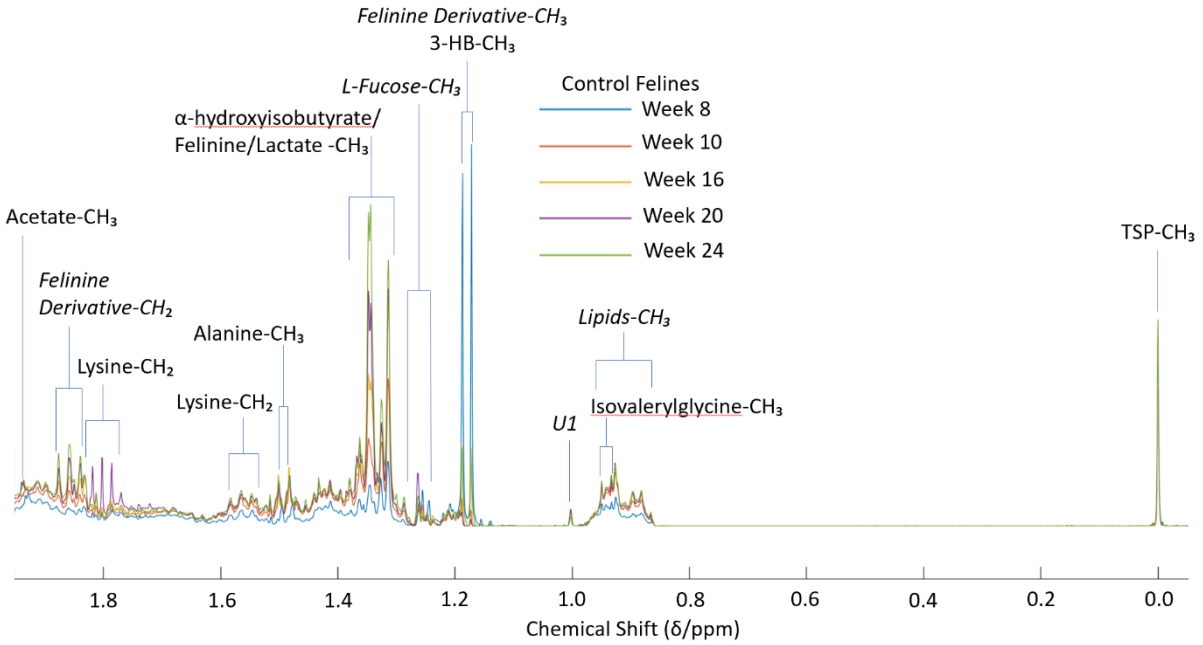
AZI is barely metabolised with no active metabolites and is predominantly excreted in bile and faeces [258]. AZI has been recovered by 6% - 12% in urine previously [258]. Therefore, unsurprisingly this drug is not observed in the ^{1}H NMR feline urinary profiles.

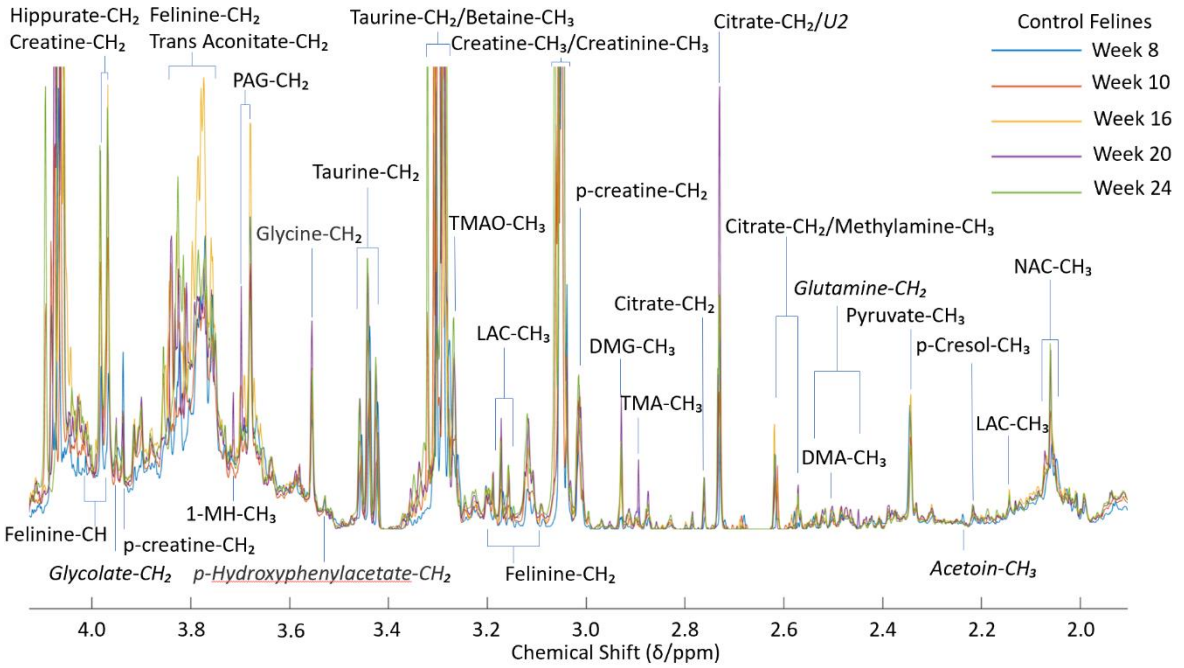
In contrast, CLAV and AMOX are metabolised unchanged in urine. CLX administration is dependent on weight of the feline and has a half-life of 1-2 hours [259]. In felines, the maximum concentration of amoxicillin and clavulanate found in plasma is 6.6µg/ml and 3.7µg/ml respectively [259]. Additionally, elimination of CLX is fast for amoxicillin and clavulanate, with 12% and 17% eliminated in the urine respectively [259]. Therefore unsurprisingly, CLX was not found in the feline urine samples. ^{1}H NMR profiles have been previously observed of clavulanate in aqueous solutions [260] and resonances for this molecule include δ = 4.78, 5.67, 3,46, 2.95, 4.89, 4.18 and 4.12 ppm [260]. Amoxicillin can be observed in human urinary profiles and distinct resonances can be observed at δ = 1.43, 1.44, 4.46, 5.16, 5.51, 5.56, 6.98, 7.37 and 9.11 ppm [261]. Amoxicilloic acid and amoxicillin diketopiperazine-2',5'-dione, metabolites of AMOX, have been observed in pigs plasma and the former, amoxicilloic acid, observe in pig urine [262]. In rats and dogs it was found that CLAV produced a metabolite 1-amino-4-hydroxybutan-2-one in addition to being excreted unchanged [263] however, in humans another metabolite was found namely, 2,5-dihydro-4-(2-hydroxyethyl)-5-oxo-1H-pyrrole-3-carboxylic acid [264]. However, these metabolites were not found in the feline urine samples. AMOX and CLAV have been shown to be unstable in different aqueous solutions at a range of pH's however at a neutral pH they are typically stable [265]. Higher acidity or alkalinity showed a significant decomposition rate in both CLAV and AMOX [265]. As the feline urine was typically a neutral pH this should not affected the decomposition of the drugs in the urinary profiles.

MET is predominantly metabolised unchanged. However, suggested metabolites have been proposed through multiple routes unlike the other drugs in this study for example, methyl-5-nitrodiazole-1-acetic acid in humans [266]. In humans and murine models, a range of metabolites have been shown to be excreted in urine including 2-carboxylic, N-(2-hydroxyethyl)-oxamic acid, acetamide, 2-hydroxymethyl, 1-acetic acid 2-methyl and glucuronide derivatives [267]. Despite this, the primary drug resonances itself which can be

found on human metabolome database [148] are not found within the ^{1}H NMR feline urinary profiles. MET has distinct ^{1}H NMR signals at δ = 8.028 and 2.463 ppm [148] which are not seen in the profiles from the felines herein.

Appendix 9 – Average Spectral Differences in Feline Urinary Profiles showing Age and Sex Groupings





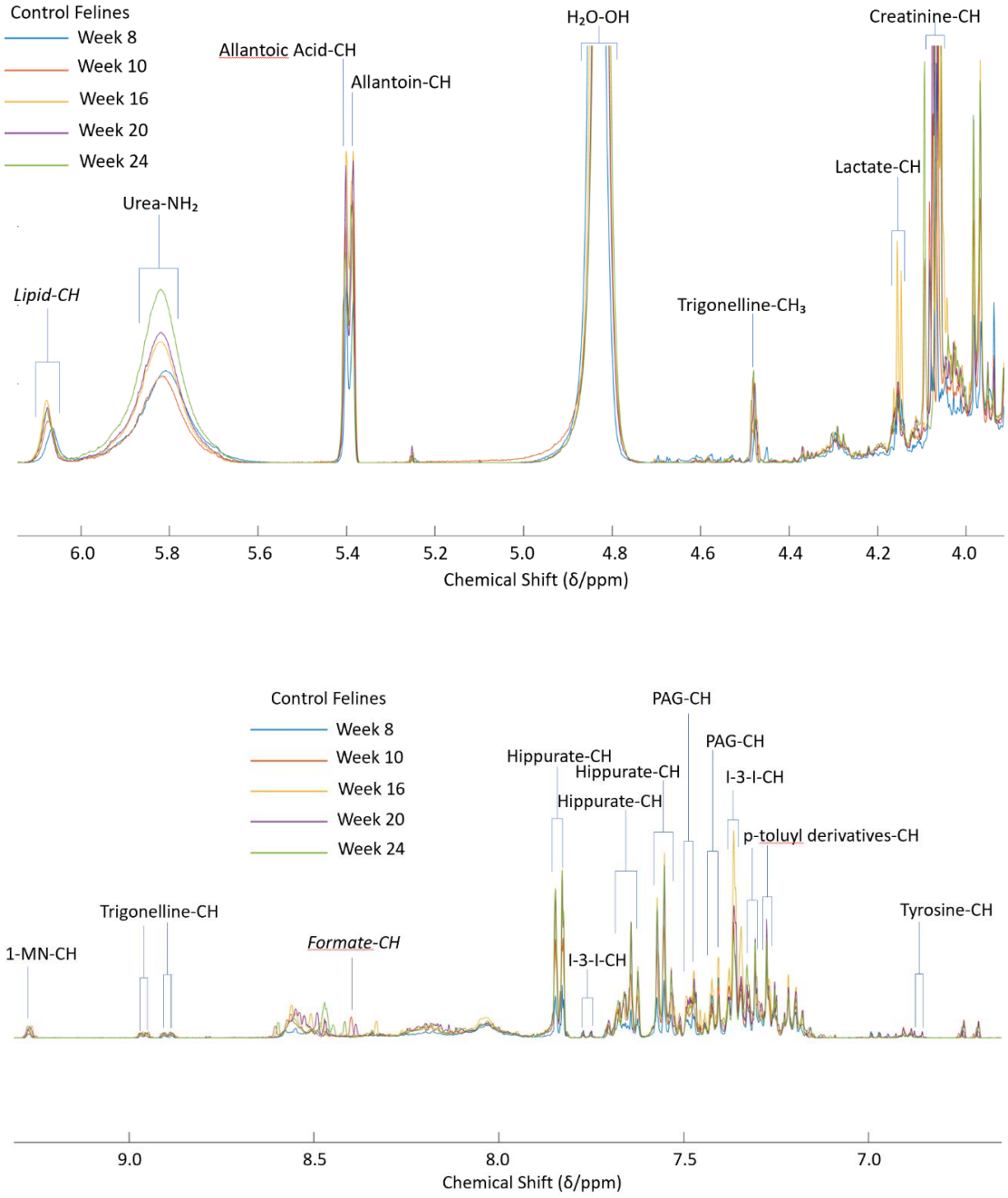
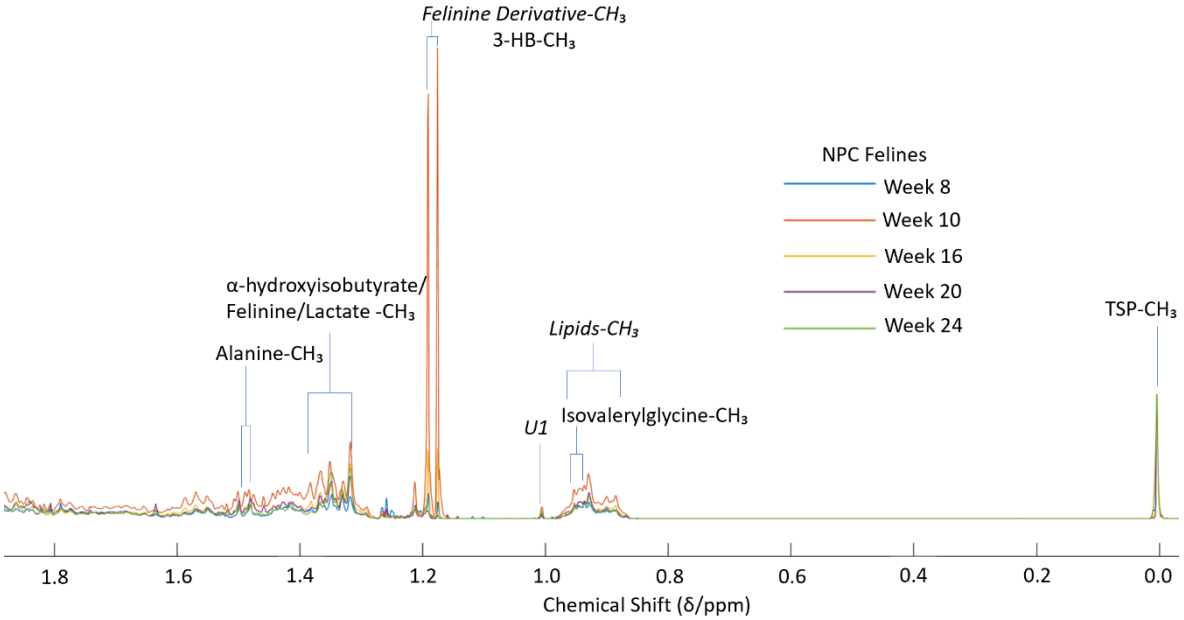
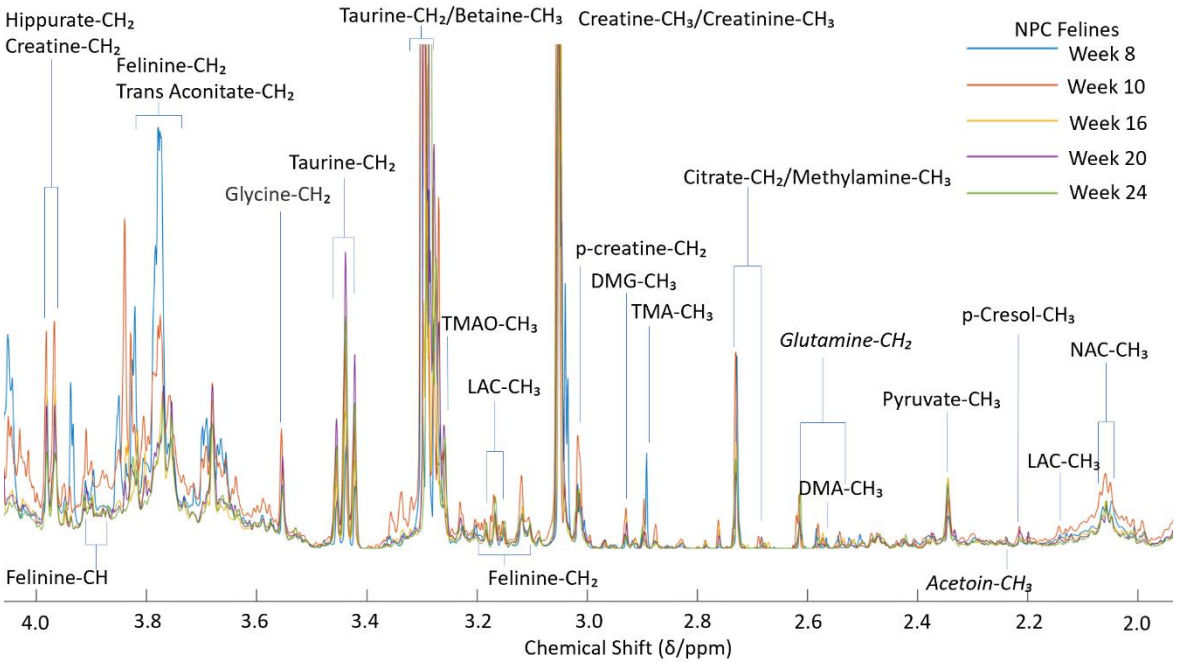


Figure 9.0. ^{1}H NMR profiles of control felines. Average stack plots of each age group at 8, 10, 16, 20 and 24 weeks in blue, orange, yellow, purple and green respectively. Formate is italicised here as there is a clear shift in the formate signal attributable to the molecule which typically arises at δ = 8.45 ppm. The other italicised molecules have not been shown to be present in whole feline urine in NMR elsewhere however, there is evidence that they are present in urine [209]. U1 and U2 represent unknown metabolites. Abbreviations: 1-MH, 1-methylhisitidine; 1-MN, 1-methylnicotinamide; 3-HB, 3-hydroxybutyrate; DMA, dimethylamine; DMG, dimethylglycine; I-3-I, indole-3-lactate; IS, Indoxyl sulphate; LAC, L-acetyl carnitine; NAC, N-acetyl groups; PAG, phenyl-acetyl-glycine; p-creatine, phosphocreatine; TMA, trimethyl amine; TMAO, trimethylamine-n-oxide


Felinine Derivative-CH3
3-HB-CH3
NPC Felines
Week 8
Week 10
Week 16
Week 20
Week 24
α-hydroxyisobutyrate/
Felinine/Lactate -CH3
Alanine-CH3
Lipids-CH3
U1
Isovalerylglycine-CH3
TSP-CH3
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Chemical Shift (δ/ppm)


Hippurate-CH2
Creatine-CH2
Felinine-CH2
Trans Aconitate-CH2
Taurine-CH2/Betaine-CH3
Creatine-CH3/Creatinine-CH3
NPC Felines
Week 8
Week 10
Week 16
Week 20
Week 24
Taurine-CH2
Glycine-CH2
Citrate-CH2/Methylamine-CH3
p-creatine-CH2
DMG-CH3
TMA-CH3
TMAO-CH3
LAC-CH3
Glutamine-CH2
p-Cresol-CH3
NAC-CH3
Pyruvate-CH3
LAC-CH3
DMA-CH3
Felinine-CH
Felinine-CH2
Acetoin-CH3
4.0
3.8
3.6
3.4
3.2
3.0
2.8
2.6
2.4
2.2
2.0
Chemical Shift (δ/ppm)

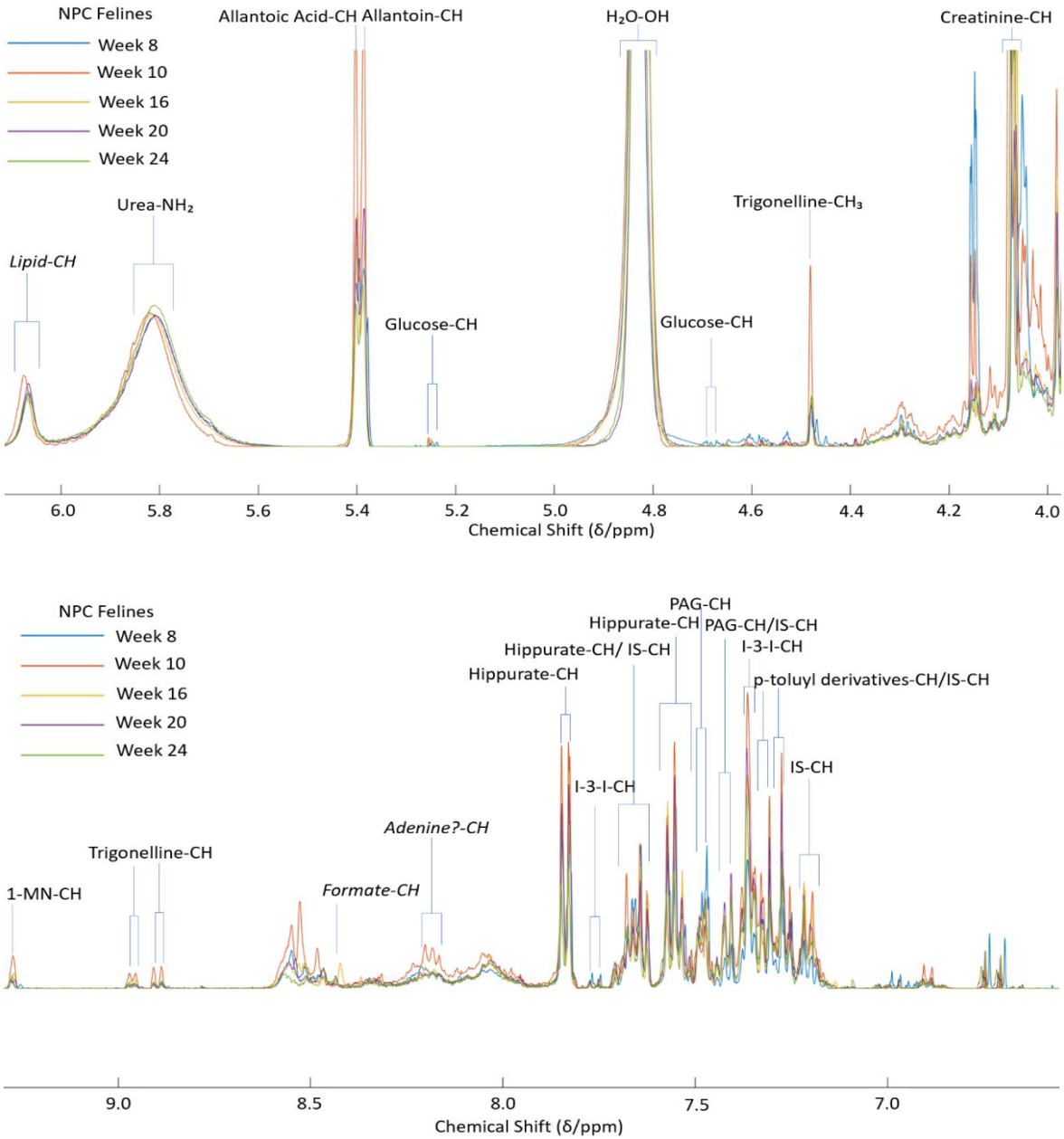


Figure 9.1. ^{1}H NMR profiles of NPC felines. Average stack plots of each age group at 8, 10, 16, 20 and 24 weeks in blue, orange, yellow, purple and green respectively. Formate is italicised here as there is a clear shift in the formate signal attributable to the molecule which typically arises at δ = 8.45 ppm. The other italicised molecules have not been shown to be present in whole feline urine in NMR elsewhere however, there is evidence that they are present in urine [209]. U1 and U2 represent unknown metabolites. Abbreviations: 1-MH, 1-methylhisitidine; 1-MN, 1-methylnicotinamide; 3-HB, 3-hydroxybutyrate; DMA, dimethylamine; DMG, dimethylglycine; I-3-I, indole-3-lactate; IS, Indoxyl sulphate; LAC, L-acetyl carnitine; NAC, N-acetyl groups; PAG, phenyl-acetyl-glycine; p-creatine, phosphocreatine; TMA, trimethyl amine; TMAO, trimethylamine-n-oxide

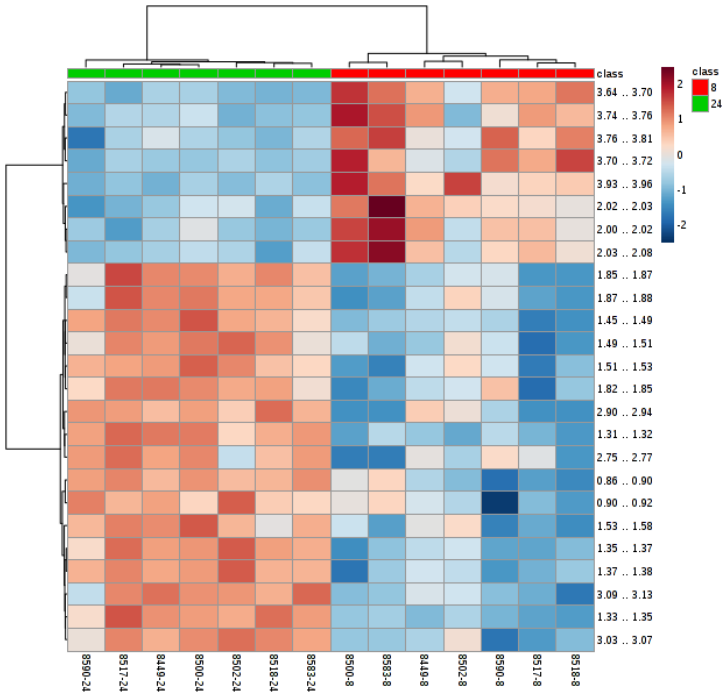


Figure 9.2. Heatmap of 7 control felines urinary profiles at 8 weeks and at 24 weeks inclusive of the top 25 metabolites

Significant bucket (ppm)	Assignment	Week 8	Week 24
3.64-3.70	Phenylacetylglycine-C<u>H</u>$_2$	↑	↓
3.74-3.76	1-Methylhisitidine-C<u>H</u>$_3$	↑	↓
3.76-3.81	Felinine-β-C<u>H</u>$_2$	↑	↓
3.70-3.72	Phenylacetylglycine-C<u>H</u>$_2$	↑	↓
3.93-3.96	Creatine-C<u>H</u>$_2$ /p-creatine-C<u>H</u>$_2$ /Glycolate-C<u>H</u>$_2$	↑	↓
2.02-2.03	N-acetyl-C<u>H</u>$_3$	↑	↓
2.00-2.02	N-acetyl-C<u>H</u>$_3$	↑	↓
2.03-2.08	N-acetyl-C<u>H</u>$_3$	↑	↓
1.85-1.87	Felinine Derivative-C<u>H</u>$_2$	↓	↑
1.87-1.88	Felinine Derivative-C<u>H</u>$_2$	↓	↑
1.45-1.49	Alanine-C<u>H</u>$_3$	↓	↑
1.49-1.51	Lysine-C<u>H</u>$_3$	↓	↑
1.51-1.53	Lysine-C<u>H</u>$_3$	↓	↑
1.82-1.85	Felinine Derivative-C<u>H</u>$_2$	↓	↑
2.90-2.94	Dimethylglycine-C<u>H</u>$_3$	↓	↑
1.31-1.32	Felinine-C<u>H</u>$_3$/L-Lactate-C<u>H</u>$_3$	↓	↑
2.75-2.77	Citrate-C<u>H</u>$_2$/DMA-C<u>H</u>$_3$	↓	↑
0.86-0.90	Unassigned-C<u>H</u>$_3$	↓	↑
0.90-0.92	Unassigned-C<u>H</u>$_3$	↓	↓
1.53-1.58	Lysine-C<u>H</u>$_3$	↓	↑
1.35-1.37	Felinine-C<u>H</u>$_3$/2-Hydroxyisobutyrate-C<u>H</u>$_3$	↓	↑
3.09-3.13	Felinine-C<u>H</u>$_2$	↓	↑
3.03-3.07	Creatine/Creatinine-C<u>H</u>$_2$	↓	↑

Table 9.0: Significant Metabolites Assigned in control felines profiles at 8 weeks and at 24 weeks

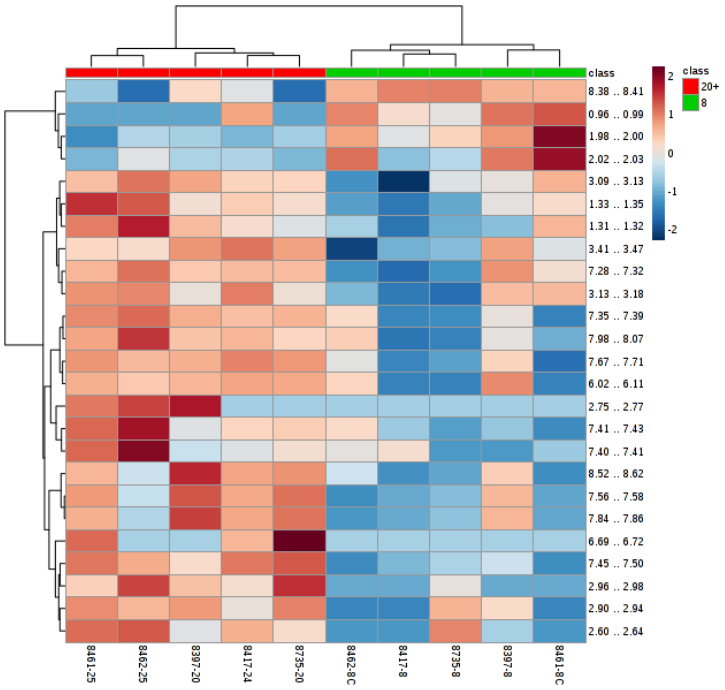


Figure 9.3. Heatmap of 5 NPC Felines urinary profiles at 8 weeks and at 20+ weeks inclusive of the top 25 metabolites

Significant bucket (ppm)	Assignment	Week 8	Week 20+
8.38-8.41	Tentative aromatic amino acid-C<u>H</u>	↑	↓
0.96-0.99	Isovarylglycine-C<u>H</u>$_3$/Lipids-C<u>H</u>$_3$	↑	↓
1.98-2.00	N-acetyl-C<u>H</u>$_3$	↑	↓
2.02-2.03	N-acetyl-C<u>H</u>$_3$	↑	↓
3.09-3.13	Felinine -C<u>H</u>$_2$	↓	↑
1.33-1.35	Felinine-C<u>H</u>$_3$/2-Hydroxyisobutyrate-C<u>H</u>$_3$	↓	↑
1.31-1.32	Felinine-C<u>H</u>$_3$/L-Lactate-C<u>H</u>$_3$	↓	↑
3.41-3.47	Taurine-C<u>H</u>$_2$-SO$_3$-	↓	↑
7.28-7.32	p-toluyl derivative-C<u>H</u>	↓	↑
3.13-3.18	L-acetyl carnitine-C<u>H</u>$_2$/Felinine-C<u>H</u>$_2$	↓	↑
7.35-7.39	Indole-3-lactate-C<u>H</u>	↓	↑
7.98-8.07	Tentative broad amino acid protein-C<u>H</u>	↓	↑
7.67-7.71	Hippurate-γ/δ/ε-C<u>H</u>	↓	↑
6.02-6.11	Lipid resonance-C<u>H</u>	↓	↑
2.75-2.77	Citrate-C<u>H</u>$_2$/DMA-C<u>H</u>$_3$	↓	↑
7.41-7.43	phenylacetylglycine-C<u>H</u>	↓	↑
8.52-8.62	Broad amino acid protein-C<u>H</u>	↓	↑
7.56-7.58	Hippurate-γ/ε-C<u>H</u>	↓	↑
7.84-7.86	Hippurate-β/ζ-C<u>H</u>	↓	↑
6.69-6.72	Unassigned	↓	↑
7.45-7.50	phenylacetylglycine-C<u>H</u>	↓	↑
2.96-2.98	Unassigned	↓	↑
2.90-2.94	Dimethylglycine-C<u>H</u>$_3$	↓	↑
2.60-2.64	Citrate-C<u>H</u>$_2$	↓	↑

Table 9.1. Significant Metabolites Assigned in control felines profiles at 8 weeks and at 24 weeks

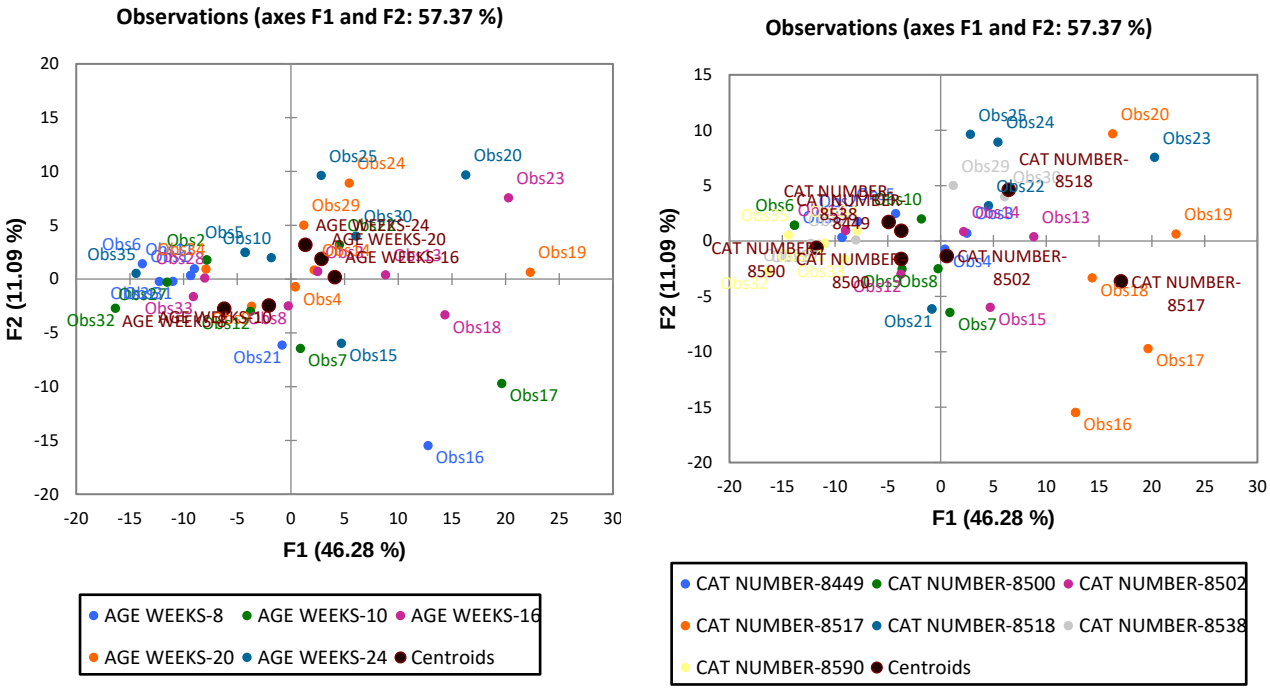


Figure 9.4. Left: PCA plot of control cat ages whereby blue, green, pink, orange and navy represent week 8, 10, 16, 20 and 24 weeks respectively. **Right:** PCA plot of control cat numbers whereby blue, green, pink, orange, navy, grey and yellow represent cat numbers 8449, 8500, 8502, 8517, 8518, 8538 and 8590 respectively.

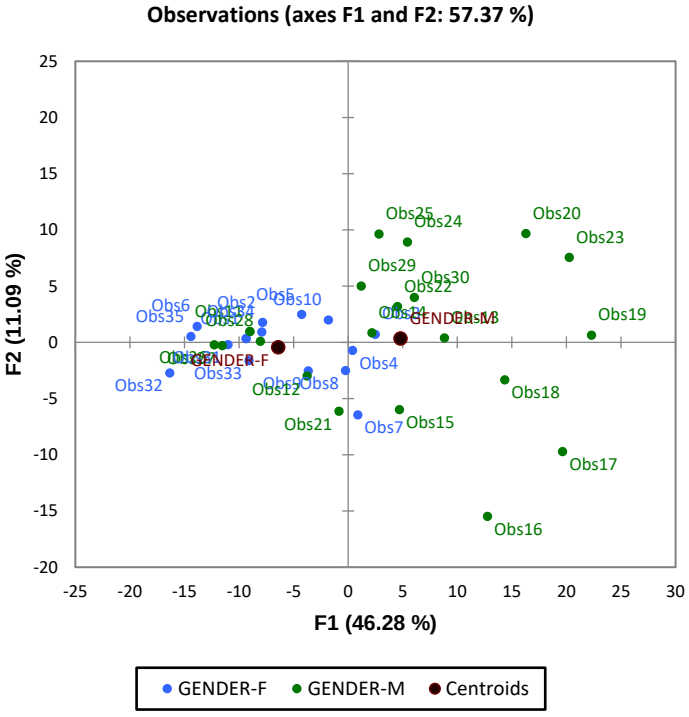


Figure 9.5. PCA plot of gender in control felines whereby female and male felines are represented in blue and green respectively.

Appendix 10 – Metabolite Confirmation using 2D NMR for Feline Urine Samples

Felinine ((R)-2-amino-3-(4-hydroxy-2-methylbutan-2-ylthio) propanoic acid) has previously been elucidated in feline urine [209] and has been found and confirmed here using 2D-TOCSY. This molecule enables an easy distinction between human and feline urine.

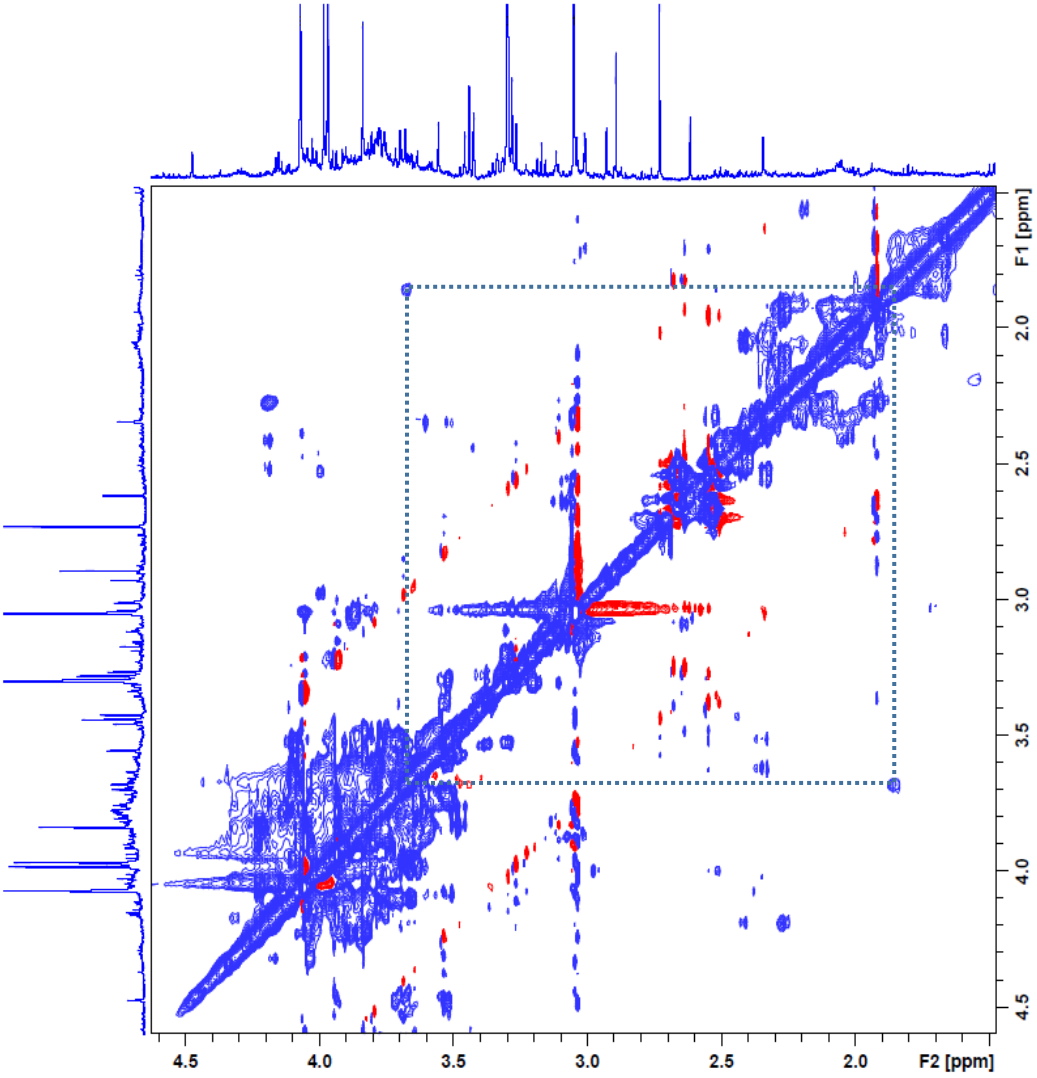


Figure 10.0. Confirmation of felinine using 400 MHz NMR 2D TOCSY; 6-C$\underline{H_2}$ (δ = 1.86 ppm) triplet and 7-C$\underline{H_2}$ (δ = 3.77 ppm) triplet.

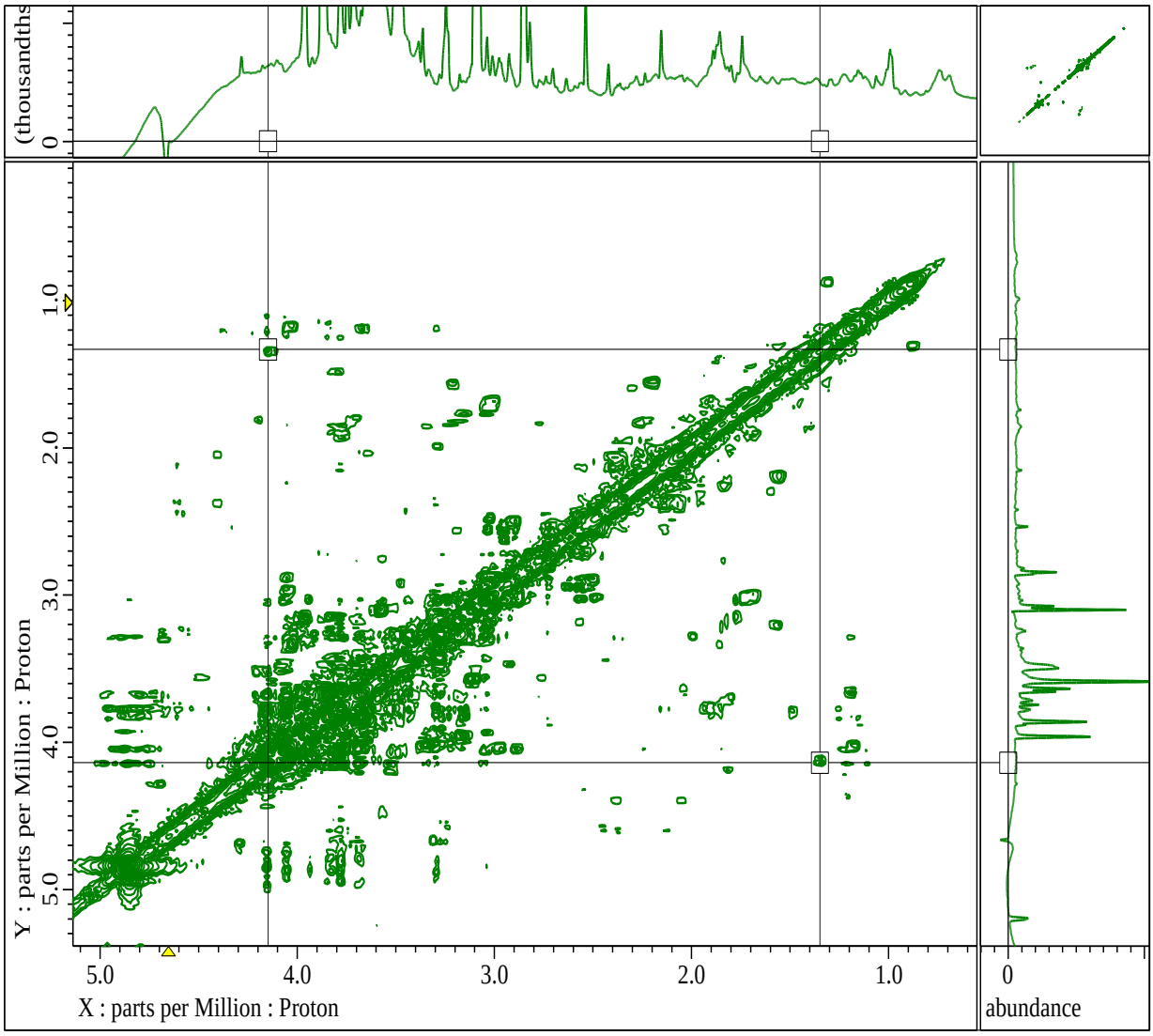


Figure 10.1. Confirmation of lactate using 600 MHz NMR 2D COSY; -C<u>H</u>$_3$ (δ = 1.33 ppm) doublet and -C<u>H</u>$_2$ (δ = 4.14 ppm) quartet. Other correlations can be observed here at the δ = 4.14 ppm resonance.

Appendix 11 – Partial 600 MHz 1D Feline Urinary Profile using WET solvent suppression sequence

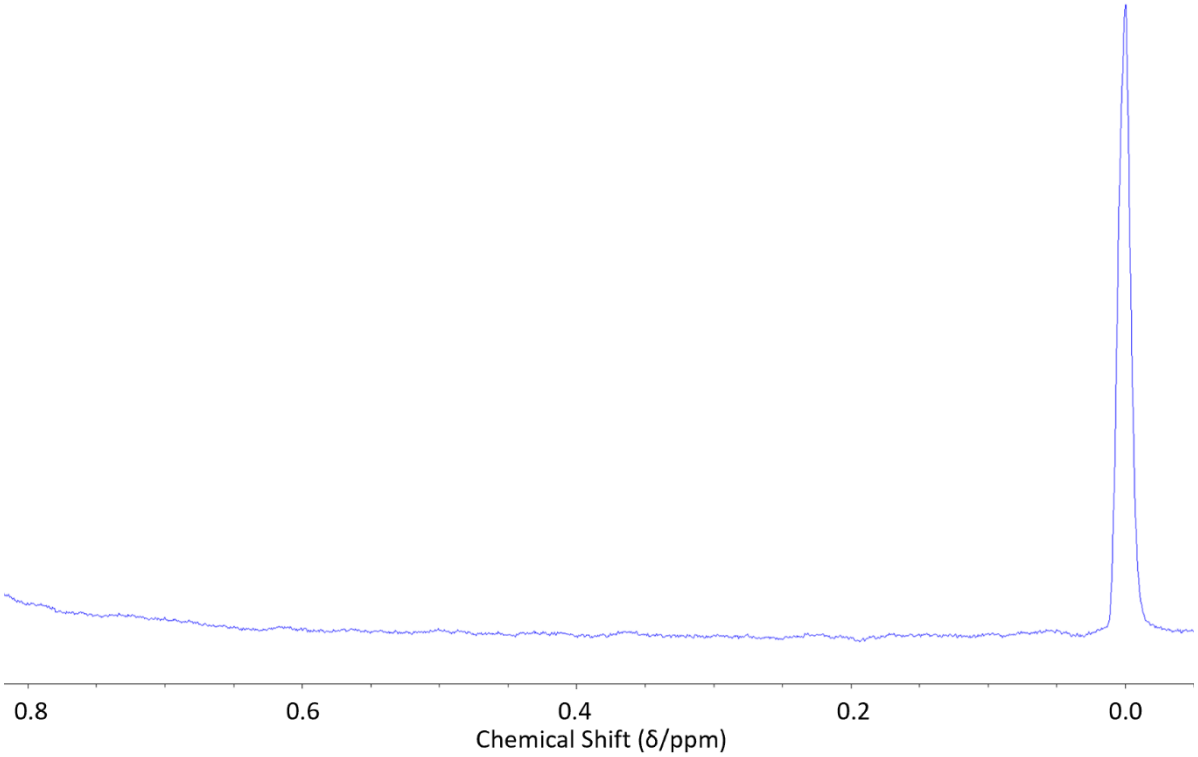


Figure 11.0. Partial 600 MHz of urinary profile from a typical feline urine sample with NPC.

600 MHz NMR spectra were acquired in order to distinguish any bile acid (BA) signals in the feline urine as a result of upregulated cholesterol. From profiles acquired herein, no BA signals were observed.

Appendix 12 – Control and Treated Felines Plasma analysis

ANOVA was performed on plasma biomarkers in the same felines within 16-50-week period. Although the dataset is substantially smaller, due to missing the earlier and later time points, a clear distinction can be observed when using an adjust *p* value of $p>0.0005$ using Fishers/LSD non parametric testing.

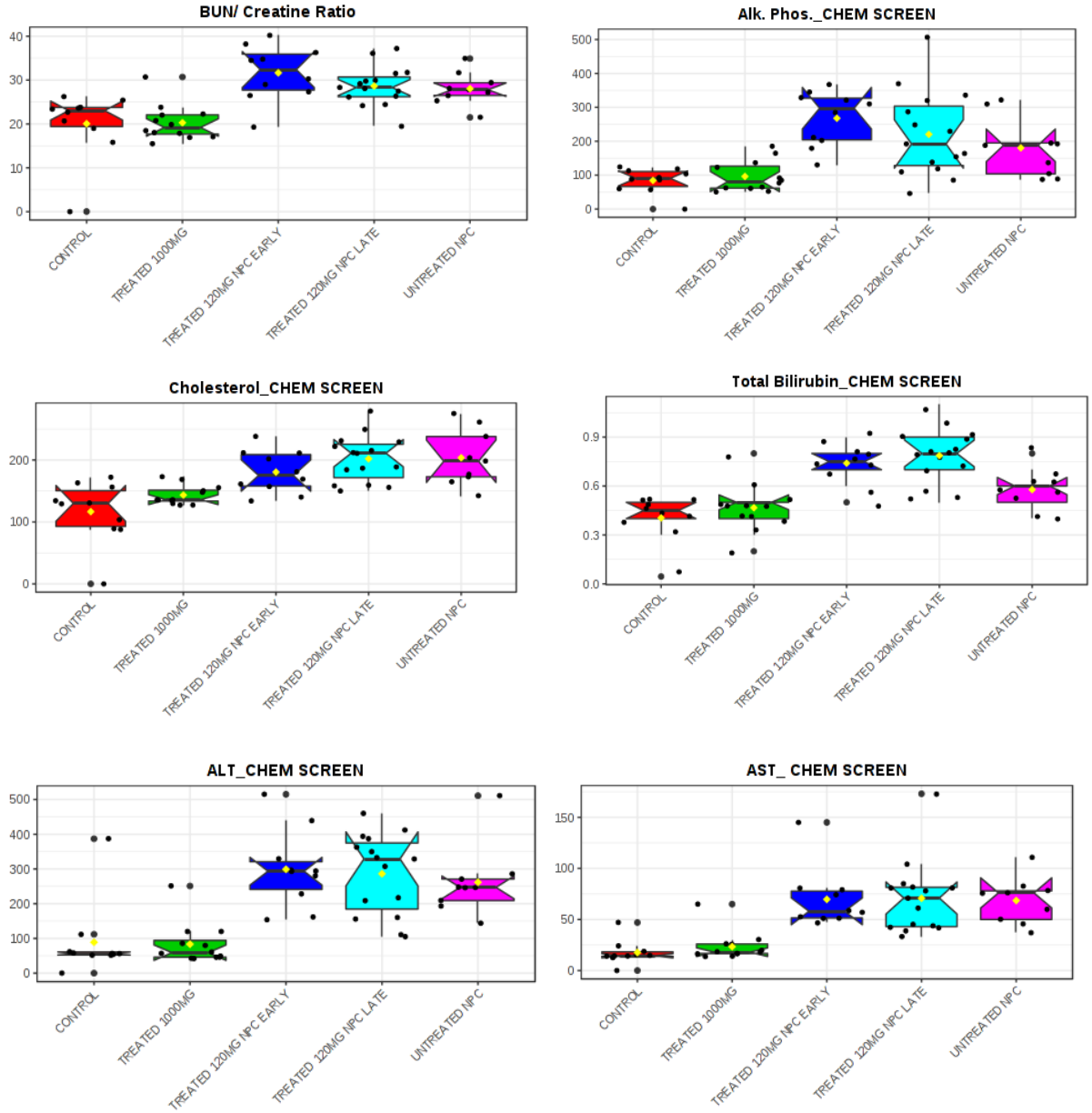


Figure 12.0. ANOVA of plasma with adjusted *p* value (FDR) cut off $p>0.0005$ using Fishers/LSDs non-parametric. Alanine transferase (ALT), Aspartate transaminase (AST), Cholesterol, Blood Urea Nitrogen (BUN)/Creatine Ratio, Alkaline Phosphatase (Alk. Phosp) and Total Bilirubin were measured in plasma from the same felines however, only aged matched plasma was available from felines >16 weeks.

Although it is unknown in the plasma samples whether changes are able to observed pre-symptomatically because of the samples collected were < 16 weeks, evident changes are observable post-symptomatically (>16 weeks).